# Guide to Climatological Practices

2018 edition

WEATHER CLIMATE WATER

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 100

# Guide to Climatological Practices

2018 edition

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 100

**EDITORIAL NOTE**

METEOTERM, the WMO terminology database, may be consulted at http://public.wmo.int/en/resources/meteoterm.

Readers who copy hyperlinks by selecting them in the text should be aware that additional spaces may appear immediately following http://, https://, ftp://, mailto:, and after slashes (/), dashes (-), periods (.) and unbroken sequences of characters (letters and numbers). These spaces should be removed from the pasted URL. The correct URL is displayed when hovering over the link or when clicking on the link and then copying it from the browser.

WMO-No. 100



Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 84 03
Fax: +41 (0) 22 730 81 17
Email: publications@wmo.int

ISBN 978-92-63-10100-6

## PUBLICATION REVISION TRACK RECORD

| *Date* | *Part/chapter/ section* | *Purpose of amendment* | *Proposed by* | *Approved by* |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

# CONTENTS

## FOREWORD

One of the purposes of the World Meteorological Organization (WMO), as laid down in the WMO Convention, is to promote the standardization of meteorological and related observations, including those that are applied to climatological studies and practices. With this aim, the World Meteorological Congress adopts from time to time Technical Regulations that lay down the meteorological practices and procedures to be followed by the Organization's Members. The Technical Regulations are supplemented by a number of Guides, which describe in more detail the practices, procedures and specifications that Members are expected to follow or implement in establishing and conducting their arrangements in compliance with the Technical Regulations and in otherwise developing their meteorological and climatological services. One of the publications in this series is the *Guide to Climatological Practices* (WMO-No. 100), whose aim is to provide, in a convenient form for all concerned with the practice of climatology, information that is of the greatest importance for the successful implementation of their work. A complete description of the theoretical bases and range of applications of climatological methods and techniques is beyond the scope of this Guide, although references to such documentation are provided wherever applicable.

The first edition of the *Guide to Climatological Practices* was published in 1960 on the basis of material developed by the Commission for Climatology; it was edited by a special working group with the assistance of the Secretariat. The second edition of the Guide originated at the sixth session of the Commission for Special Applications of Meteorology and Climatology (1973). The Commission instructed the working group responsible for the Guide to make arrangements for a substantially revised edition in the light of the progress made in climatology during the preceding decade and in the use of climatological information and knowledge in various areas of meteorology and other disciplines. The seventh session of the Commission (1978) re-established the working group, which continued to work on the second edition on a chapter-by-chapter basis, producing the version that was ultimately published in 1983. Since then, climate-related activities have expanded in virtually every area of human life, particularly in science and public policy. The *Guide to Climatological Practices* is a key resource that is designed to help Members provide a seamless stream of crucial information for daily practices and operations in National Meteorological and Hydrological Services (NMHSs).

The work on the third edition of the Guide started in 1990 when the content and authorship were approved by the advisory working group of the Commission for Climatology at a meeting in Norrköping, Sweden. An Editorial Board for the Guide was subsequently established to supervise individual lead authors and chapter editors. Nevertheless, it was not until 1999 that the lead authors received a draft summary to further develop the text of the Guide. In the following year, the Editorial Board met in Reading, United Kingdom, and defined the content and further details of each chapter. In 2001, the Commission for Climatology, at its thirteenth session, decided to establish an Expert Team on the *Guide to Climatological Practices* with clear terms of reference to expedite the process. While Part I of the publication had been made available on the Web, a major effort was needed to finalize Part II and the presentation of information on specialized requirements for the provision of climate services. The fourteenth session of the Commission, in 2005, re-established the Expert Team and agreed that some overarching activities would be the responsibility of the Management Group. Those included the further development of Part II of the Guide and more work on the review and designation of Regional Climate Centres (RCCs). The Expert Team met in Toulouse, France, that year, and decided to compile a full, integrated draft text, including annexes, of the third edition of the Guide.

With the collective effort and expertise provided by a large number of authors, editors and internal and external reviewers, the text of the third edition of the Guide was finally approved by the President of the Commission for Climatology just before the fifteenth session of the Commission, held in Antalya, Turkey, in February 2010. The Commission, at its sixteenth session held in Heidelberg, Germany, in July 2014, decided to establish a Task Team on the *Guide to Climatological Practices* for the regular updating of this publication to ensure the timely availability of the best scientific knowledge to users. The current version of the Guide, which was approved by the Commission for Climatology at its seventeenth session, held in Geneva, in April 2018, is the result of the work of this Task Team.

This updated version of the Guide is published in English and will be gradually translated in the other official languages of WMO to maximize dissemination of knowledge. As with the previous versions, WMO Members may translate this Guide into their national languages, as appropriate.

I wish to express my gratitude to the WMO Commission for Climatology for taking the initiative to oversee this long process. On behalf of the World Meteorological Organization, I also wish to thank all those who have contributed to the preparation of this publication. Special recognition is due to Dr Thomas C. Peterson, the outgoing President of the Commission for Climatology (2010–2018), who guided and supervised the preparation of the text during the fifteenth and sixteenth intersessional periods of the Commission. I also wish to acknowledge the significant contributions of Dr Govindarajalu Srinivasan (India), Lead of the Task Team on the *Guide to Climatological Practices,* and Dr Nathaniel Guttman (United States of America), lead scientific editor of this current version of the publication.

(Petteri Taalas)
Secretary-General

# CHAPTER 1. INTRODUCTION

## 1.1 PURPOSE AND CONTENT OF THE GUIDE

This publication is designed to provide Members of the World Meteorological Organization (WMO) with guidance and assistance in developing national activities linked to climate information and services. There have been two previous editions of the Guide: the original publication, which appeared in 1960, and the second edition, which was published in 1983. While many fundamentals of climate science and climatological practices have remained constant over time, scientific advances in climatological knowledge and data analysis techniques, as well as changes in technology, computer capabilities and instrumentation, have made the second edition obsolete.

The third edition describes essential basic principles and modern practices in the development and implementation of all climate services, and outlines best practices in climatology. It presents concepts and considerations, and provides references to other technical guidance and information sources, rather than attempting to be all-inclusive in the guidance presented. The current version was published in 2011, however, it will be regularly reviewed to keep its content up to date.

This first chapter includes information on climatology and its scope, the organization and functions of a national climate service, and international climate programmes. The remainder of the Guide is broken down into six chapters (Climate observations, stations and networks; Climate data management; Characterizing climate from datasets; Statistical methods for analysing datasets; Climate products and their dissemination; Service delivery) and one annex (International climate activities).

The procedures described in the Guide have been taken, where possible, from decisions on standards and recommended practices and procedures. The main decisions concerning climate practices are contained in the WMO Technical Regulations and Manuals, and reports of the World Meteorological Congress and the Executive Council, and originate mainly from recommendations of the Commission for Climatology. Lists of relevant WMO and other publications of particular interest to those working in climatology are provided in the reference section at the end of each chapter.

## 1.2 CLIMATOLOGY

Climatology is the study of climate, the workings of the climate system, its variations and extremes and its influences on a variety of activities including, but far from limited to, water resources, human health, safety and welfare. Climate, in a narrow sense, can be defined as the average weather conditions for a particular location and period of time. Climate can be described in terms of statistical descriptions of the central tendencies and variability of relevant elements such as temperature, precipitation, atmospheric pressure, humidity and winds, or through combinations of elements, such as weather types and phenomena, that are typical of a location or region, or of the earth as a whole, for any time period. Beyond being treated as a statistical entity, climate can also be studied as a determinant of, a resource for and a hazard to human activities.

Climatology considers both climatic variability and change. Climatic variability refers to variations in climate conditions from time period to time period (e.g. intra-seasonal, inter-annual and inter-decadal). In general, climatic variability is connected with variations in the state of the atmospheric and ocean circulation and land surface properties (e.g. soil moisture) at the intra-seasonal to inter-decadal timescales. Climate change, in contrast, refers to a systematic change in the statistical properties of climate (e.g. mean and variance) over a prolonged period (e.g. decades to centuries) as manifested in an upward or downward trend in, for example, extreme rainfall values. For most of the Earth's climate history, systematic changes of climate have occurred because of natural causes such as variations in the nature of the Earth's orbit around the sun or solar output and the changing relationship between the "natural" components that make up the climate system (see 1.2.2). However, there is now mounting evidence that humans and their activities constitute an important component of the climate system.

### 1.2.1 History

Early references to the weather can be found in the poems of ancient Greece and in the Old Testament of the Judeo-Christian Bible. Even older references appear in the Vedas, the most ancient Hindu scriptures, which were written about 1800 B.C. Specific writings on the theme of meteorology and climatology are found in Hippocrates' *On Airs, Waters and Places*, dated around 400 B.C., followed by Aristotle's *Meteorologica*, written around 350 B.C. To the early Greek philosophers, climate meant "slope" and referred to the curvature of the Earth's surface, which gives rise to the variation of climate with latitude due to the changing incidence of the Sun's rays. Logical and reliable inferences on climate are to be found in the work of the Alexandrian philosophers Eratosthenes and Aristarchus.

With the onset of extensive geographical exploration in the fifteenth century, descriptions of the Earth's climates and the conditions giving rise to them started to emerge. The invention of meteorological instruments such as the thermometer in 1593 by Galileo Galilei and the barometer in 1643 by Evangelista Torricelli gave a greater impulse to the establishment of mathematical and physical relationships between the different characteristics of the atmosphere. This in turn led to the establishment of relationships that could describe the state of the climate at different times and in different places.

The observed pattern of circulation linking the tropics and subtropics, including the trade winds, tropical convection and subtropical deserts, was first interpreted by George Hadley in 1735, and subsequently became known as the Hadley cell. Julius von Hann, who published the first of three volumes of the *Handbook of Climatology* in 1883, wrote the classic work on general and regional climatology, which included data and eyewitness descriptions of weather and climate. In 1918, Wladimir Köppen produced the first detailed classification of world climates based on the vegetative cover of land. This endeavour was followed by more detailed developments in descriptive climatology. The geographer E.E. Federov (1927), for example, attempted to describe local climates in terms of daily weather observations. In the first thirty years of the twentieth century, the diligent and combined use of global observations and mathematical theory to describe the atmosphere led to the identification of large-scale atmospheric patterns. Notable in this field was Sir Gilbert Walker, who conducted detailed studies of the Indian monsoon, the Southern Oscillation, the North Atlantic Oscillation and the North Pacific Oscillation.

Other major works on climatology included those by Tor Bergeron (on dynamic climatology, published in 1930) and Wladimir Köppen and Rudolf Geiger, who produced a climatology handbook in 1936. Geiger first described the concept of microclimatology in some detail in 1927, but the development of this field did not occur until the Second World War. During the war, for planning purposes, a probability risk concept of weather data for months or even years ahead was found to be necessary and was tried. C.W. Thornthwaite established a climate classification in 1948 based on a water budget and evapotranspiration. In the following decades the development of theories on climatology saw major progress.

In September 1929, the International Meteorological Organization (IMO) Conference of Directors, meeting in Copenhagen, agreed to set up a technical commission for climatology "for the study of all questions relating to this branch of science". The creation of WMO in 1950 (as the successor to IMO, which was founded in 1873) established a system of data collection and led to the systematic analysis of climate and to conclusions about the nature of climate. During the latter decades of the twentieth century, the issue of climate change drew attention to the need to understand climate as a major part of a global system of interacting processes involving all of the Earth's major domains (see 1.2.2). Climate change is defined as a statistically significant alteration in either the average state of the climate or in its variability, persisting for an extended period, typically decades or longer. It may be caused by natural internal processes, external forcing or persistent anthropogenic (resulting from or produced by human activity) changes in the composition of the atmosphere or in land use. Considerable national and international efforts are being directed at other aspects of climatology as well. These efforts include improved measurement and monitoring of climate, increased understanding of the causes and patterns of natural variability, more reliable methods of predicting climate over seasons and years ahead, and better understanding of the linkages between climate and a range of social and economic activities and ecological changes.

Recent climate-related activities have been mainly shaped and impacted by three world climate conferences, which WMO organized in collaboration with other United Nations agencies and programmes, in 1979, 1990 and 2009. More details on these three conferences are mentioned in Annex1, International climate activities.

### 1.2.2 The climate system

The climate system (Figure 1.1) is a complex, interactive system consisting of the atmosphere, land surface, snow and ice, oceans and other bodies of water, and living organisms. The atmosphere is the gaseous envelope surrounding the Earth. The dry atmosphere consists almost entirely of nitrogen and oxygen, but also contains small quantities of argon, helium, carbon dioxide, ozone, methane and many other trace gases. The atmosphere also contains water vapour, condensed water droplets in the form of clouds, and aerosols. The hydrosphere is that part of the Earth's climate system comprising liquid water distributed on and beneath the Earth's surface in oceans, seas, rivers, freshwater lakes, underground reservoirs and other water bodies. The cryosphere collectively describes elements of the Earth system containing water in its frozen state and includes all snow and ice (sea ice, lake and river ice, snow cover, solid precipitation, glaciers, ice caps, ice sheets, permafrost and seasonally frozen ground). The surface lithosphere is the upper layer of the solid Earth, including both the continental crust and the ocean floor. The biosphere comprises all ecosystems and living organisms in the atmosphere, on land (terrestrial biosphere) and in the oceans (marine biosphere), including derived dead organic matter, such as litter, soil organic matter and oceanic detritus.

Influenced by solar radiation and the radiative properties of the Earth's surface, the climate of the Earth is determined by interactions among the components of the climate system. The interaction of the atmosphere with the other components plays a dominant role in forming the

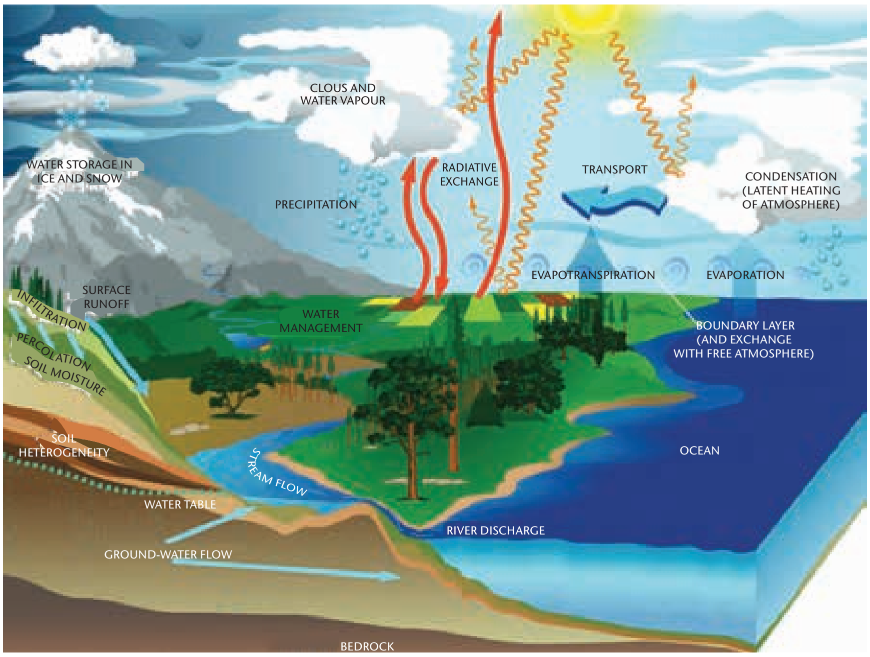



**Figure 1.1 The climate system**

climate. The atmosphere obtains energy directly from incident solar radiation or indirectly via processes involving the Earth's surface. This energy is continuously redistributed vertically and horizontally through thermodynamic processes or large-scale motions with the unattainable aim of achieving a stable and balanced state of the system. Water vapour plays a significant role in the vertical redistribution of heat through condensation and latent heat transport. The ocean, with its vast heat capacity, limits the rate of temperature change in the atmosphere and supplies it with water vapour and heat. The distribution of the continents affects oceanic currents, and mountains redirect atmospheric motions. The polar, mountain and sea ice reflects solar radiation back into space. In high latitudes sea ice acts as an insulator and protects the ocean from rapid energy loss to the much colder atmosphere. The biosphere, including its human activities, affects atmospheric components such as carbon dioxide, as well as features of the Earth's surface such as soil moisture and albedo.

Interactions among the components occur on all scales (Figures 1.2 and 1.3). Spatially, the microscale encompasses climate characteristics over small areas such as individual buildings and plants or fields. A change in microclimate can be of major importance when the physical characteristics of an area change. New buildings may produce extra windiness, reduced ventilation, excessive runoff of rainwater, and increased pollution and heat. Natural variations in microclimate, such as those related to shelter and exposure, sunshine and shade, are also important: they can determine, for example, which plants will prosper in a particular location or the need to provide for safe operational work and leisure activities. The mesoscale encompasses the climate of a region of limited extent, such as a river catchment area, valley, conurbation or forest. Mesoscale variations are important in applications including land use, irrigation and damming, the location of natural energy facilities, and resort location. The macroscale encompasses the climate of large geographical areas, continents and the globe. It determines national resources and constraints in agricultural production and water management, and is thus linked to the nature and scope of human health and welfare. It also defines and determines the impact of major features of the global circulation such as the El Niño/Southern Oscillation (ENSO), the monsoons and the North Atlantic Oscillation.

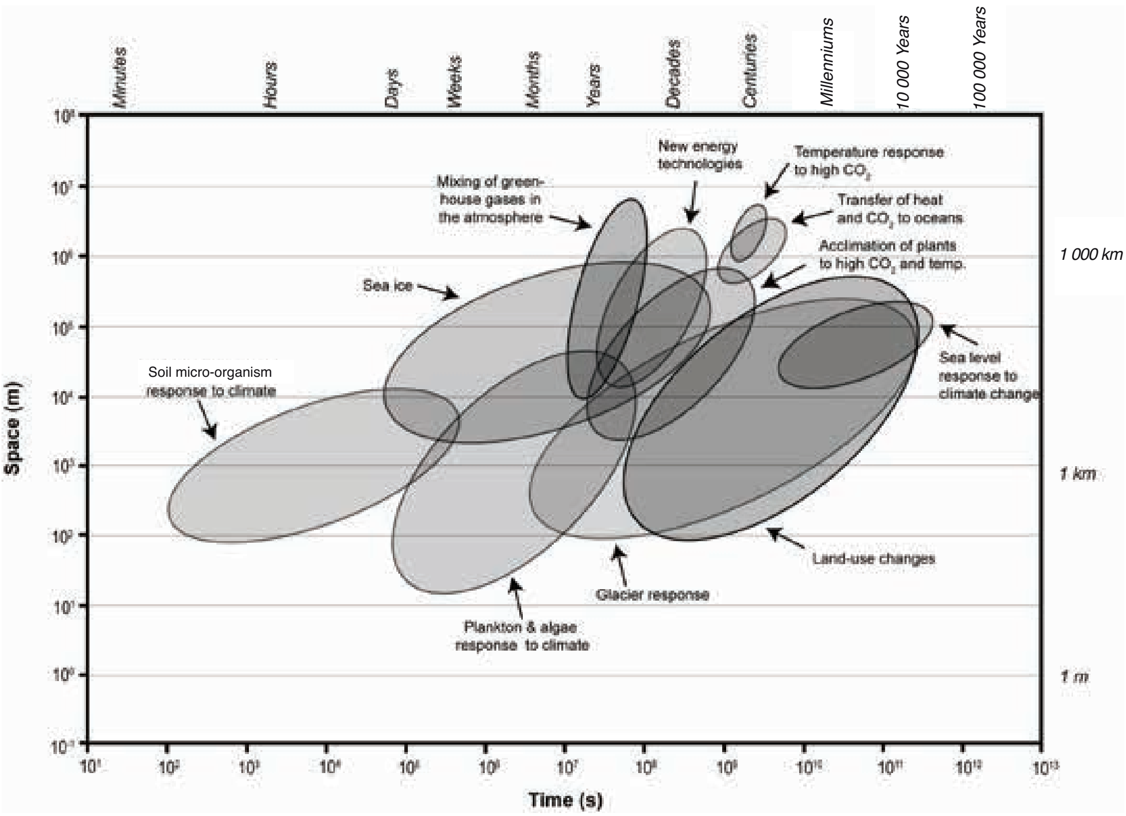



**Figure 1.2 Temporal and spatial scales (courtesy of Todd Albert, United States)**

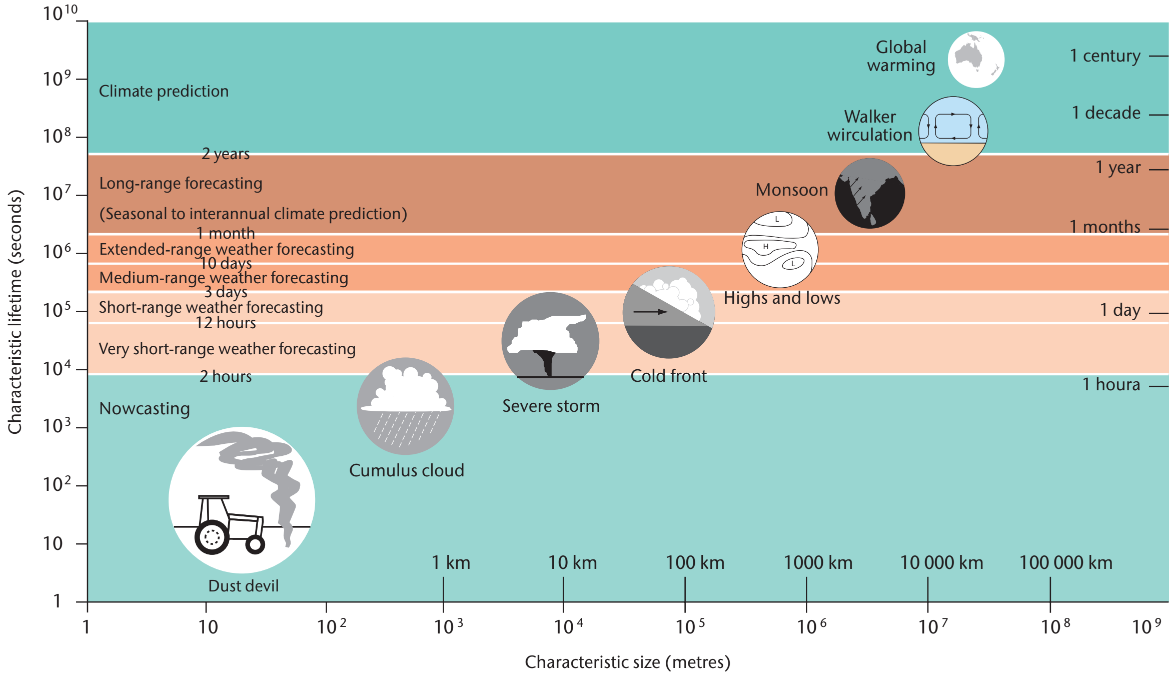



**Figure 1.3. Lifetime of atmospheric phenomena**

*Source:* J.W. Zillman, WMO Bulletin, Volume 48(2), 1999

A temporal scale is an interval of time. It can range from minutes and hours to decades, centuries and longer. The characteristics of an environmental element over an hour are important, for example, in agricultural operations such as pesticide control and in monitoring energy usage for heating and cooling. The characteristics of an element over a day might determine the human activities that can be safely pursued. The climate over months or years will determine, for example, the crops that can be grown or the availability of drinking water and food. Longer timescales of decades and centuries are important for studies of climate variation caused by natural phenomena such as atmospheric and oceanic circulation changes and by the activities of humans.

Climate change has become a major issue for the human community. Human activities, especially the burning of fossil fuels, have led to changes in the composition of the global atmosphere. Marked increases in tropospheric carbon dioxide and methane during the industrial era, along with increased aerosols and particulate emissions, are significantly affecting global climate. Chlorofluorocarbons extensively used in the past as aerosol propellants, cleaning fluids and refrigerants are a main cause of stratospheric ozone depletion. Over one fifth of the world's tropical forests was cleared between 1960 and 2000, likely altering the complex mesoscale and global hydrological cycles. Artificial canyons formed in cities by buildings, together with asphalt road surfaces, increase the amount of radiation absorbed from the Sun and form urban heat islands. Accelerated runoff of rainwater and the removal of trees and other vegetation reduce the amount of transpired water vapour that would otherwise help to moderate temperature. Pollution from vehicles and buildings accumulates, especially in calm conditions, and causes many human health problems and damage to structures.

Conscious of the growing worldwide concern about the danger of irreversible changes occurring in the natural environment, WMO has taken the lead in promoting studies of changes in the climate system and their effects on mankind, on world energy and food production and on water reserves. Climate change and its potential consequences, as well as those effects that have already occurred, have become key topics occupying decision-makers in recent years, and in some countries these concerns are second only to economic and defence matters. Even in these two areas, climate is involved in strategic planning and tactical decision-making. Many international conferences have been held to devise ways of reducing the human impact on climate and to design strategies to exploit climate for social and economic benefit. These meetings include the World Climate Conferences in Geneva in 1979, 1990 and 2009; the United Nations Conference on Environment and Development in Rio de Janeiro in 1992; and the World Summit on Sustainable Development in Johannesburg in 2002. The establishment of the Intergovernmental Panel on Climate Change (IPCC) and the Intergovernmental Negotiating Committee for a Framework Convention on Climate Change also marked important milestones in addressing changes in climate that are related to human activities.

### 1.2.3 **Uses of climatological information and research**

Climatology has become a dynamic branch of science with a broad range of functions and applications. New techniques are being developed and investigations are being undertaken to study the application of climate in many fields, including agriculture, forestry, ecosystems, energy, industry, production and distribution of goods, engineering design and construction, human well-being, transportation, tourism, insurance, water resources and disaster management, fisheries, and coastal development. Viable research programmes on the climate system and its broad influence, and on the applications of climate knowledge for societal benefits are needed to enable climatologists to inform and advise users and answer a myriad of questions about the climate. Previously, the study of climate provided basic data, information and techniques to delineate local, mesoscale and global climates. While these are primary deliverables, they are also the raw material for deeper analysis and for services when coupled with other social, economic and physical data. The crucial roles of climate data and climate predictions in planning for disaster mitigation and sustainable development, and in addressing all the consequences of climate change, are now firmly established within various conventions, such as the United Nations Framework Convention on Climate Change.

Applied climatology makes maximum use of meteorological and climatological knowledge and information for solving practical social, economic and environmental problems. Climatological services are designed for a variety of public, commercial and industrial users. Assessments of the effects of climate variability and climate change on human activities, as well as the effects of human activities on climate, are major factors in local, national and global economic development, social programmes and resource management.

Current interest in the impact of economic development and other human activities on climate and how climate variability and change influence human societies highlights the need for further research into the physical and dynamical processes involved in the climate system, as well as the need for their statistical description. Understanding of natural climate variability, appreciation of climate sensitivity to human activities, and insight into the predictability of weather and climate for periods ranging from days to decades are fundamental to improving our ability to respond effectively to economic and societal problems. Physical climatology embraces a wide range of studies that include interactive processes of the climate system. Dynamic climatology is closely related to physical climatology, but it is mainly concerned with patterns in the general circulation of the atmosphere. Both involve the description and study of the properties and behaviour of the atmosphere.

Improving the prediction of climate is now a substantial global activity. Initially, predictions were based on empirical and statistical techniques, but now they derive more and more from expanded numerical weather prediction techniques. Increasingly complex models that represent and couple together the atmosphere, oceans, land interface, sea ice, and atmospheric aerosols and gases are being developed. The models can be used to simulate climate change over several decades and also to predict seasonal or inter-annual variations in climate. Such seasonal outlooks generally take the form of the probability that the value of an element, such as the mean temperature or aggregated rainfall over a period, will be above, near or below normal. Seasonal outlooks presently show skill in prediction for regions where there is a strong relationship between sea-surface temperature and weather, such as in many tropical areas. Because of their probabilistic nature, however, much care is needed in their dissemination and application. Decision-making that incorporates climate information is a growing area of investigation.

All climate products and services, from information derived from past climate and weather data to estimations of future climate, for use in research, operations, commerce and government, are underpinned by data obtained by extensively and systematically observing and recording a number of key variables that enable the characterization of climate on a wide range of timescales. The adequacy of a climate service is highly dependent on the spatial density and accuracy of the observations and on the data management processes. Without systematic observations of the climate system, there can be no climate services.

The need for more accurate and timely information continues to increase rapidly as the diversity of users' requirements continues to expand. It is in the interest of every country to apply consistent practices in performing climate observations, in handling climate records and in maintaining the necessary quality and usefulness of the services provided.

## 1.3 INTERNATIONAL CLIMATE PROGRAMMES

The WMO Commission for Climatology (CCl) is concerned with the overall requirements of WMO Members for advice, support and coordination in many climate activities. The Commission has been known by slightly different names and has seen its terms of reference change in accordance with changing demands and priorities, but it has effectively been in operation since it was established in 1929 under the International Meteorological Organization. It provides overall guidance for the implementation of the World Climate Programme within WMO. Additional details about international climate programmes are contained in Annex 1.

## 1.4 GLOBAL AND REGIONAL CLIMATE ACTIVITIES

All countries should understand and provide for the climate-related information needs of the public. This requires climate observations, management and transmission of data, various data services, climate system monitoring, practical applications and services for different user groups, forecasts on sub-seasonal and inter-annual timescales, climate projections, policy-relevant assessments of climate variability and change, and research priorities that increase the potential benefits of all these activities. Many countries, especially developing and least developed countries, may not have sufficient individual capacity to perform all of these services. The World Climate Conference-3, held in Geneva in 2009, proposed the creation of a Global Framework for Climate Services to strengthen the production, availability, delivery and application of science-based climate prediction and services. The Framework is intended to provide a mechanism for developers and providers of climate information, as well as climate-sensitive sectors around the world, to work together to help the global community better adapt to the challenges of climate variability and change.

The World Meteorological Organization has developed a network of Global Producing Centres for Long-Range Forecasts ( GPCLRFs) and Regional Climate Centres (RCCs) to help Members cope effectively with their climate information needs. The definitions and mandatory functions of GPCLRFs and RCCs are contained in the *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485), Part I, and are part of the WMO Technical Regulations. Part II of the Manual also provides the criteria for the designation of GPCLRFs, RCCs and other operational centres by WMO.

The designated GPCLRFs produce global long-range forecasts according to the criteria defined in the *Manual on the Global Data-processing and Forecasting System* and are recognized by WMO on the basis of the recommendation of the Commission for Basic Systems. In addition, WMO has established a Lead Centre for Long-range Forecast Multi-model Ensembles and the Lead Centre for the Long-range Forecast Verification System, which provide added value to the operational services of GPCLRFs.

The Regional Climate Centres are designed to help WMO Members in a given region to deliver better and more consistent climate services and products, such as regional long-range forecasts, and to strengthen the capability of Members to meet national climate information needs. The primary clients of an RCC are National Meteorological and Hydrological Services (NMHSs) and other RCCs in the region and in neighbouring areas. The services and products from the RCCs are provided to the NMHSs for further definition and dissemination and are not distributed to users without the permission of the NMHSs within the region. The responsibilities of an RCC do not duplicate or replace those of NMHSs. It is important to note that NMHSs retain the mandate and authority to provide the liaison with national user groups and to issue advisories and warnings, and that all RCCs are required to adhere to the principles of WMO Resolution 40 (Cg-XII) concerning the exchange of data and products.

The complete suite of products and services of RCCs can vary from one region to another, on the basis of the priorities established by the relevant Regional Association. There will, however, be certain essential functions all WMO-designated RCCs must perform to established criteria, thus ensuring a certain uniformity of RCC services around the globe. These functions include:

(a) Operational activities for long-range forecasting, including interpretation and assessment of relevant outputs from GPCLRFs, generation of regional and sub-regional tailored products, and preparation of consensus statements concerning regional or subregional forecasts;

(b) Climate monitoring, including regional and subregional climate diagnostics, analysis of climate variability and extremes, and implementation of regional climate watches for extreme climate events;

(c) Data services to support long-range forecasting, including the development of regional climate datasets;

(d) Training in the use of operational RCC products and services.

In addition to these mandatory RCC functions, a number of activities are highly recommended. These include downscaling of climate change scenarios, non-operational data services such as data rescue and data homogenization, coordination functions, training and capacity-building, and research and development.

Recognizing that climate information can be of substantial benefit in adapting to and mitigating the impacts of climate variability and change, WMO has helped establish Regional Climate Outlook Forums. Using a predominantly consensus-based approach, the forums have an overarching responsibility to produce and disseminate a regional assessment of the state of the regional climate for the upcoming season. The forums bring together national, regional and international climate experts on an operational basis to produce regional climate outlooks based on input from NMHSs, regional institutions, RCCs and GPCLRFs. They facilitate enhanced feedback from the users to climate scientists, and catalyse the development of user-specific products. They also review impediments to the use of climate information, share successful lessons regarding applications of past products and enhance sector-specific applications. The regional forums often lead to national forums developing detailed national-scale climate outlooks and risk information, including warnings, for decision-makers and the public.

The Regional Climate Outlook Forum process, which can vary in format from region to region, typically includes at least the first of the following activities and, in some instances, all four:

(a) Meetings of regional and international climate experts to develop a consensus for regional climate outlooks, usually in a probabilistic form;

(b) A broader forum involving both climate scientists and representatives from user sectors, for the presentation of consensus climate outlooks, discussion and identification of expected sectoral impacts and implications, and the formulation of response strategies;

(c) Training workshops on seasonal climate prediction to strengthen the capacity of national and regional climate scientists;

(d) Special outreach sessions involving media experts to develop effective communication strategies.

## 1.5 NATIONAL CLIMATE ACTIVITIES

In most countries, NMHSs have long held key responsibilities for national climate activities, including the making, quality control and storage of climate observations; the provision of climatological information; research on climate; climate prediction; and the applications of climate knowledge. There has been, however, an increasing contribution to these activities from academia and private enterprise.

Some countries have within their NMHSs a single division responsible for all climatological activities. In other countries, the NMHS may find it beneficial to assign responsibilities for different climatological activities (such as observation, data management and research) to different units within the Service. The division of responsibilities could be made on the basis of commonality of skills, such as across synoptic analysis and climate observation, or across research in weather and climate prediction. Some countries establish area or branch offices to handle subnational activities, while in other cases the necessary pooling and retention of skills for some activities are achieved through a regional cooperative entity serving the needs of a group of countries.

When there is a division of responsibility within an NMHS, or in those cases when responsibilities are assigned to another institution altogether, it is essential that a close liaison exist between those applying the climatological data in research or services and those responsible for the acquisition and management of the observations. This liaison is of paramount importance in determining the adequacy of the networks and of the content and quality control of the observations. It is also essential that the personnel receive training appropriate to their duties, so that the climatological aspects are handled as effectively as would be the case in an integrated climate centre or division. If data are handled in several places, it is important to establish a single coordinating authority to ensure that there is no divergence among datasets. Climatologists within an NMHS should be directly accountable for, or provide consultation and advice regarding:

(a) Planning of station networks;

(b) Location or relocation of climatological stations;

(c) Care and security of the observing sites;

(d) Regular inspection of stations;

(e) Selection and training of observers;

(f) Instruments or observing systems to be installed so as to ensure that representative and homogeneous records are obtained (see Chapter 2).

Once observational data are acquired, they must be managed. Functions involved in the management of information from observing sites include data and metadata acquisition, quality control, storage, archiving and access (see Chapter 3). Dissemination of the collected climatic information is another requirement. An NMHS must be able to anticipate, investigate and understand the needs for climatological information among government departments, research institutions and academia, commerce, industry and the general public; promote and market the use of the information; make available its expertise to interpret the data; and advise on the use of the data (see Chapters 6 and 7).

An NMHS should maintain a continuing research and development programme or establish a working relationship with an institution that has research and development capabilities directly related to the climatological functions and operations of the NMHS. The research programme should consider new climate applications and products that increase user understanding and application of climate information.

Studies should explore new and more efficient methods of managing an ever-increasing volume of data, improving user access to the archived data, and migrating data to digital form. Quality assurance programmes for observations and summaries should be routinely evaluated with the goal of developing better and timelier techniques. The use of information dissemination platforms such as the Internet should also be developed.

The fulfilment of national and international responsibilities, and the building of NMHS capacity relevant to climate activities, can be achieved only if adequately trained personnel are available. Thus, an NMHS should maintain and develop links with training and research establishments dealing with climatology and its applications. In particular, it should ensure that personnel attend training programmes that supplement general meteorological training with education and skills specific to climatology. The WMO Education and Training Programme fosters and supports international collaboration that includes the development of a range of mechanisms for continued training, such as fellowships, conferences, familiarization visits, computer-assisted learning, training courses and technology transfer to developing countries. In addition, other WMO Programmes, such as the World Climate Programme, the Hydrology and Water Resources Programme and the Agricultural Meteorology Programme, undertake capacity-building activities relevant to climate data, monitoring, prediction, applications and services. To be successful, a national climate service programme must have a structure that works effectively within a particular country. The structure must be one that allows the linkage of available applications, scientific research, technological capabilities and communications into a unified system. The essential components of a national climate service programme are:

(a) Mechanisms to ensure that the climate information and prediction needs of all users are recognized;

(b) Collection of meteorological and related observations, management of databases and provision of data;

(c) Coordination of meteorological, oceanographic, hydrological and related scientific research to improve climate services;

(d) Multidisciplinary studies to determine national risk and sectoral and community vulnerability related to climate variability and change, to formulate appropriate response strategies, and to recommend national policies;

(e) Development and provision of climate information and prediction services to meet user needs;

(f) Linkages to other programmes with similar or related objectives to avoid unnecessary duplication of efforts.

It is important to realize that a national climate service programme constitutes an ongoing process that may change in structure over time. An integral part of this process is the continual review and incorporation of user requirements and feedback in order to develop useful products and services. Gathering requirements and specifications is vital in the process of programme development. Users may contribute through the evaluation of products, which invariably leads to refinement and the development of improved products. Measuring the benefits of the application of products can be a difficult task, but interaction with users through workshops, training and other outreach activities will aid the process. The justification for a national climate service programme, or requests for international financial support for aspects of the programme, can be greatly strengthened by well-documented user requirements and positive feedback. The documented endorsement of the programme, by one or more representative sections of the user community, is essential to guide future operations and to assist in the promotion of the service as a successful entity.

## 1.6 QUALITY MANAGEMENT

One of the ways to ensure an effective national climate service programme is to implement a quality management system (QMS). Many NMHSs strive for quality observations and ultimately the provision of quality products and services. Thus the concept of quality is not a foreign one but how that translates into a system may be unfamiliar. By definition a QMS is a set of policies, processes and procedures required for the planning and execution of various core business processes to improve the performance of the NMHSs. A QMS enables the organizations to identify, measure, control and improve the various core business processes that will ultimately lead to improved business performance and enhanced customer satisfaction.

The components of a QMS are determined by the NMHS and its objectives and goals. A QMS based on an international standard such as that established by ISO is helpful as it provides a set of requirements that can be adopted in order to drive quality within NMHSs. It also provides for an external auditing of the processes and thus gives an independent validation of the system. The adoption of the ISO 9001:2015 quality standard will ensure a consciously established quality; this is preferable to hoping it will evolve by chance.

When developing a quality management system based on ISO 9001:2015, the following quality management principles should be taken into account:

- Customer focus;
- Leadership;
- Engagement of people;
- Process approach;
- Improvement;
- Evidence-based decision making;
- Relationship management.

The approach underpinning the ISO 9001:2015 quality management standard is based on the management of processes and their interaction using what is known as the Plan-Do-Check-Act (PDCA) cycle. This involves the establishment of objectives and processes while considering resources on hand to meet customer requirements. What was planned is then implemented and the resulting outputs, such as products and services, are checked to see whether they are in line with the plan. Should improvements be identified these are then dealt with and the whole cycle is repeated. During the cycle there is a need to address any risks and opportunities that may be identified to increase the efficiency of the system.

The benefits of a QMS are:

- Consistency in the issuing of quality products and services;
- Assistance in meeting customer needs;
- More efficient ways of working that will save time, money and resources;
- Improved operational performance by reducing errors;
- Motivation and engagement of staff with more efficient internal processes;
- Facilitation and identification of training opportunities;
- Helping to identify risk and opportunities.

A QMS helps coordinate and direct an organization’s activities to meet customer and regulatory requirements and improve its effectiveness and efficiency on a continuous basis. It should thus be an end-to-end QMS covering all activities from raw measurements and observations to services delivered to end users. Such QMS seeks to improve quality and performance in order to meet or exceed customer expectations all the while taking into account the NMHSs context as well as interested parties' expectations and requirements. A quality management system is thus an important part of the climatological practises of NMHSs and plays a key role in driving quality through the whole value chain, from the setting up of instrumentation, through the storage and quality control of data, to the production of climatological products and services. This value chain is made stronger by the people involved; while no one person is involved in every aspect of the chain, their competencies are necessary to achieve the set objectives (*Technical Regulations* (WMO-No. 49), Volume I).

Thus the provision of climate services by NMHSs can be improved by the implementation of a quality management system, and it is recommended that ISO 9001:2015 be used to achieve this. National Meteorological and Hydrological Services will be able to identify and meet the requirements of their clients thanks to the comprehensive set of WMO Technical Regulations and guidance documents, which provide a sound foundation for the operation of NMHSs and compliance with national and international regulatory requirements, and thanks also to the application of the principles of quality management. They will also be able to monitor and measure their own performance, to identify and mitigate risks and to find opportunities to continually improve their service delivery. In so doing, they will be in a position to apply for certification as an ISO accredited organization.

A detailed description of a quality management system and its implementation in a meteorological service can be found in *Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers* (WMO-No. 1100). This publication will be of great value when an NMHS is setting up a QMS. The *Guidelines on Quality Management in Climate Services* (WMO-No. 1221), prepared by the WMO Commission for Climatology, provide more information regarding the implementation of quality management in the context of climate services. The following Web pages may also be of assistance: http://qmc.mgm.gov.tr and http://www.bom.gov.au/wmo/quality_management/qm_publications.shtml.

## REFERENCES


Aristotle, circa 350 B.C.: *Meteorologica*.

Bergeron, T., 1930: Richtlinien einer dynamischen Klimatologie. *Meteorologische Zeitung*, 47:246–262.

Federov, E.E., 1927: Climate as totality of the weather. *Monthly Weather Review*, 55:401–403.

Geiger, R., 1927: *Das Klima der bodennahen Luftschicht. Ein Lehrbuch der Mikroklimatologie*. Fourth edition, 1961. Braunschweig, Vieweg.

Hadley, G., 1735: Concerning the cause of the general trade-winds. *Philosophical Transactions,* 39, 58-62.

Hippocrates, circa 400 B.C.: *On Airs, Waters, and Places*, available at http://classics.mit.edu//Hippocrates/airwatpl.html.

International Organization for Standardization, 2015: *ISO 9001:2015 Quality Management Systems – Requirements*.

Köppen, W., 1918: Klassification der Klimate nach Temperatur, Niederschlag und Jahreslauf. *Petermanns. Geographische Mitteilungen*, 64:193–203, 243–248.

Köppen, W. and R. Geiger , 1936: Das geographische System der Klimate, in *Handbuch der Klimatologie*, Volume I, Part C. Berlin, Borntraeger.

Thornthwaite, C.W., 1948. An Approach toward a rational classification of climate. *Geographical Review*, 38(1):55–94.

Von Hann, J., 1908: *Handbuch der Klimatologie*, three volumes. Stuttgart, Englehorn.

Walker, G.T., 1923-24: World weather, I and II. *Memoirs of the Indian Meteorological. Department,* 24(4):9.

World Meteorological Organization, 2015: *Technical Regulations* (WMO-No.49), Volume I: General Meteorological Standards and Recommended Practices, 2017 update. Geneva.

———, 2017: *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485). Geneva.

———, 2017: *Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers* (WMO-No. 1100). Geneva.


## FURTHER READING


Burroughs, W., 2003: *Climate: Into the 21st Century*. Cambridge, Cambridge University Press.

Geiger, R., R.H. Aron and P. Todhunter, 2003: *The Climate Near the Ground*. Sixth edition. Lanham, Maryland, Rowman and Littlefield Publishers.

Global Climate Observing System (GCOS), 2004: *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC* (WMO-TD No. 1219, GCOS-92). Geneva.

———, 2010: *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC* (WMO/TD-No. 1523, GCOS-138). Geneva.

Group on Earth Observations (GEO), 2005: Global Earth Observation System of Systems (GEOSS): 10-year Implementation Plan Reference Document. GEO 1000R/ESA SP-1284. Noordwijk, European Space Agency Publications Division, ESTEC.

———, 2008: *GEO 2007–2009 Work Plan - Toward Convergence*. Geneva.

Integrated Global Observing Strategy (IGOS), 2007: *Cryosphere Theme Report: For the Monitoring of our Environment from Space and from Earth* (WMO-TD No. 1405). Geneva.

Intergovernmental Panel on Climate Change (IPCC), 2004: *16 Years of Scientific Assessment in Support of the Climate Convention*. Geneva.

———, 2007: *Climate Change 2007, The Fourth Assessment Report*, Volumes 1–3. Cambridge, United Kingdom, Cambridge University Press.

Landsberg, H., 1962: *Physical Climatology*. Dubois, Pennsylvania, Gray Printing.

Mann, M.E., R.S. Bradley and M.K. Hughes, 1999: Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, 26(6):759.

McGregor, G.R., 2006: Climatology: Its scientific nature and scope. *International Journal of Climatology*, 26 (1):1-5.

McGregor, G.R., 2015: Climatology in support of climate risk management: A progress report. *Progress in Physical Geography*, 39:536-553.

World Meteorological Organization, 1986: *Report of the International Conference on the Assessment of the Role of Carbon Dioxide and of Other Greenhouse Gases in Climate Variations and Associated Impacts* (Villach, Austria, 9–15 October 1985) (WMO-No. 661). Geneva.

———, 1990: *Forty Years of Progress and Achievement: A Historical Review of WMO* (Sir Arthur Davies, ed.) (WMO-No. 721). Geneva.
———, 1990: *The WMO Achievement: 40 Years in the Service of International Meteorology and Hydrology* (WMO-No. 729). Geneva.
———, 1992: *International Meteorological Vocabulary* (WMO-No. 182). Geneva.
———, 1997: *Report of the GCOS/GOOS/GTOS Joint Data and Information Management Panel*, Third session (Tokyo, Japan, 15–18 July 1997) (WMO/TD-No. 847, GCOS- 39, GOOS- 11, GTOS- 11). Geneva.
———, 2000: *WMO – 50 Years of Service* (WMO-No. 912). Geneva.
———, 2003: *Proceedings of the Meeting on Organization and Implementation of Regional Climate Centres* (Geneva, 27–28 November 2003) (WMO/TD-No. 1198, WCASP-No. 62). Geneva.
———, 2005: *The World Climate Research Programme Strategic Framework 2005–2015 – Coordinated Observation and Prediction of the Earth System (COPES)* (WMO/TD-No. 1291, WCRP-123). Geneva.
———, 2008: *Report of the CCl/CBS Intercommission Technical Meeting on Designation of Regional Climate Centres* (Geneva, 21–22 January 2008) (WMO-TD No. 1479, WCASP-No. 77). Geneva.
———, 2009: *WMO Statement on the Status of the Global Climate in 2008* (WMO-No. 1039). Geneva.
———, 2011: *Guide to the Quality Management System for the Provision of Meteorological Service for International Air Navigation* (WMO-No.1001), 2014 update; a joint WMO/ICAO guide.
———, 2016: *Executive Council: Abridged Final Report of the Sixty-Eighth Session* (WMO-No. 1168), Resolution 5 (EC-68) - Competencies for provision of climate services. Geneva.

# CHAPTER 2. CLIMATE OBSERVATIONS, STATIONS AND NETWORKS

## 2.1 INTRODUCTION

All national climate activities, including research and applications, are primarily based on observations of the state of the atmosphere or weather. The Global Observing System provides observations of the state of the atmosphere and ocean surface. It is operated by National Meteorological and Hydrological Services (NMHS), national or international satellite agencies and several organizations and consortiums dealing with specific observing systems or geographic regions. The WMO Global Observing System is a coordinated system of different observing subsystems that provides in a cost-effective way high-quality, standardized meteorological and related environmental and geophysical observations, from all parts of the globe and outer space. Examples of the observing subsystems relevant to climate are the Global Climate Observing System (GCOS) Surface Network (GSN), the GCOS Upper-air Network (GUAN), the Global Climate Observing System (GCOS) Reference Upper-air Network (GRUAN), Regional Basic Climatological Networks, the Global Atmosphere Watch (GAW), marine observing systems, and the satellite-based Global Positioning System. The observations from these networks and stations are required for the timely preparation of weather and climate analyses, forecasts, warnings, climate services, and research for all WMO Programmes and relevant environmental programmes of other international organizations.

This chapter on observations specifies the elements needed to describe the climate and the stations at which these elements are measured, the required instrumentation, the siting of stations, network design and network operations. The guidance is based on the WMO *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), the *Guide to the Global Observing System* (WMO-No. 488), the *Guidelines on Climate Observation Networks and Systems* (WMO/TD-No. 1185) and the *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160). Each edition of the *Guide to Meteorological Instruments and Methods of Observation* has a slightly different emphasis. For example, the sixth edition (WMO, 1996) contains valuable information on sensor calibration, especially of the basic instrumentation used at climate stations, but Tables 2 and 3 of the fifth edition (WMO, 1983) provide more information about the accuracy of measurements that are needed for general climatological purposes. Cross references to other WMO publications containing more detailed guidance are provided in the sections below.

Guidance is also based on ten climate monitoring principles set forth in the *Report of the GCOS/GOOS/GTOS Joint Data and Information Management Panel* (WMO/TD-No. 847):

1. The impact of new systems or changes to existing systems should be assessed prior to implementation;

2. A suitable period of overlap for new and old observing systems is required;

3. The details and history of local conditions, instruments, operating procedures, data processing algorithms, and other factors pertinent to interpreting data (metadata) should be documented and treated with the same care as the data themselves;

4. The quality and homogeneity of data should be regularly assessed as a part of routine operations;

5. Consideration of the needs for environmental and climate monitoring products and assessments should be integrated into national, regional and global observing priorities;

6. Operation of historically uninterrupted stations and observing systems should be maintained;

7. High priority for additional observations should be given to data-poor areas, poorly observed parameters, areas sensitive to change, and key measurements with inadequate temporal resolution;

8. Long-term requirements should be specified to network designers, operators and instrument engineers at the outset of system design and implementation;

9. Enabling research observing systems to perform long-term operations in a carefully planned manner should be promoted;

10 Data management systems that facilitate access, use and interpretation of data and products should be included as essential elements of climate monitoring systems.

These principles were established primarily for surface-based observations, but they also apply to data for all data platforms. Additional principles specifically for satellite observations are listed in section 2.3.4.

## 2.2 CLIMATIC ELEMENTS

A climatic element is any one of the properties of the climate system described in section 1.2.2. Combined with other elements, these properties describe the weather or climate at a given place for a given period of time. Every meteorological element that is observed may also be regarded as a climatic element. The most commonly used elements in climatology are air temperature (including maximum and minimum), precipitation (rainfall, snowfall and all kinds of wet deposition, such as hail, dew, rime, hoar frost and precipitating fog), humidity, atmospheric motion (wind speed and direction), atmospheric pressure, evaporation, sunshine, and present weather (for example, fog, hail and thunder). Properties of the land surface and subsurface (including hydrological elements, topography, geology and vegetation), of the oceans and of the cryosphere are also used to describe climate and its variability.

The subsections below describe commonly observed elements for specific kinds of stations and networks of stations. Details can be found in the *Manual on the Global Observing System* (WMO-No. 544), the *Technical Regulations* (WMO-No. 49), in particular Volume III, and the *Guide to Agricultural Meteorological Practices* (WMO-No. 134). These documents should be kept readily available and consulted as needed.

### 2.2.1 Surface and subsurface elements

An ordinary climatological station provides the basic land area requirements for observing daily maximum and minimum temperature and amount of precipitation. A principal climatological station usually provides a broader range of observations of weather, wind, cloud characteristics, humidity, temperature, atmospheric pressure, precipitation, snow cover, sunshine and solar radiation. In order to define the climatology of precipitation, wind, or any other specific element, it is sometimes necessary to operate a station to observe one or a subset of these elements, especially where the topography is varied. Reference climatological stations (see 2.5) provide long-term, homogeneous data for the purpose of determining climatic trends. It is desirable to have a network of these stations in each country, representing key climate zones and areas of vulnerability.

In urban areas, weather can have a significant impact. Heavy rains can cause severe flooding; snow and freezing rain can disrupt transport systems; and severe storms with accompanying lightning, hail and high winds can cause power failures. High winds can also slow or stop the progress of cars, recreational vehicles, railcars, transit vehicles and trucks. The urban zone is especially susceptible to land falling tropical storms because of the large concentrations of people at risk, the high density of man-made structures, and the increased risk of flooding and contamination of potable water supplies. Urban stations usually observe the same elements as principal climatological stations, with the addition of air pollution data such as low-level ozone and other chemicals and particulate matter.

Marine observations are generally either physical-dynamical or biochemical. The physical-dynamical elements (such as wind, temperature, salinity, wind and swell waves, sea ice, ocean currents and sea level) play an active role in changing the marine system. The biochemical elements (such as dissolved oxygen, nutrients and phytoplankton biomass) are generally not active in the physical-dynamical processes, except perhaps at long timescales, and thus are called passive elements. From the perspective of most NMHSs, high priority should generally be given to the physical-dynamical elements, although in some cases biochemical elements could be important when responding to the needs of stakeholders (for example, observations related to the role of carbon dioxide in climate change).

In some NMHSs with responsibilities for monitoring hydrological events, hydrological planning, or hydrological forecasting and warning, it is necessary to observe and measure elements specific to hydrology. These elements may include combinations of river, lake and reservoir level; streamflow; sediment transport and deposition; rates of abstraction and recharge; water and snow temperatures; ice cover; chemical properties of water; evaporation; soil moisture; groundwater level; and flood extent. These elements define an integral part of the hydrologic cycle and play an important role in the variability of climate.

In addition to surface elements, subsurface elements such as soil temperature and moisture are particularly important for application to agriculture, forestry, land-use planning and land-use management. Other elements that should be measured to characterize the physical environment for agricultural applications include evaporation from soil and water surfaces, sunshine, short- and long-wave radiation, plant transpiration, runoff and water table, and weather observations (especially hail, lightning, dew and fog). Ideally, measurements of agriculturally important elements should be taken at several levels between 200 cm below the surface and 10 m above the surface. Consideration should also be given to the nature of crops and vegetation when determining the levels.

Proxy data are measurements of conditions that are indirectly related to climate, such as phenology, ice core samples, varves (annual sediment deposits), coral reefs and tree ring growth. Phenology is the study of the timing of recurring biological events in the animal and plant world, the causes of their timing with regard to biotic and abiotic forces, and the interrelation among phases of the same or different species. Leaf unfolding, flowering of plants in spring, fruit ripening, colour changing and leaf fall in autumn, as well as the appearance and departure of migrating birds, animals and insects are all examples of phenological events. Phenology is an easy and cost-effective system for the early detection of changes in the biosphere and therefore complements the instrumental measurements of national meteorological services very well.

An ice core sample contains snow and ice and trapped air bubbles. The composition of a core, especially the presence of hydrogen and oxygen isotopes, relates to the climate of the time the ice and snow were deposited. Ice cores also contain inclusions such as windblown dust, ash, bubbles of atmospheric gas, and radioactive substances in the snow deposited each year. Various measurable properties along the core profiles provide proxies for temperature, ocean volume, precipitation, chemistry and gas composition of the lower atmosphere, volcanic eruptions, solar variability, sea surface productivity, desert extent and forest fires. The thickness and content of varves are similarly related to annual or seasonal precipitation, streamflow, and temperature.

Tropical coral reefs are very sensitive to changes in climate. Growth rings relate to the temperature of the water and to the season in which the rings grew. Analysis of the growth rings can match the water temperature to an exact year and season. Data from corals are used to estimate past El Niño–Southern Oscillation (ENSO) variability, equatorial upwelling, changes in subtropical gyres, trade wind regimes and ocean salinity.

Tree-ring growth shows great inter-annual variability and also large spatial differences. Some of the variation can be related to weather and climate conditions in the microscale and macroscale; plants can be viewed as integrative measurement devices for the environment. Since trees can live for centuries, annual growth rings in some tree species can provide a long historical indication (predating instrumental measurements) of climate variability. Because of the close relationship between plant development and weather and climate, phenological observation networks are run by NMHSs in many countries.

The table below summarizes the most common surface and subsurface climatic elements that are observed for various networks or types of station.

**Examples of surface and subsurface elements for different station networks or types of stations**

| *Element* | *Ordinary climate station* | *Principal climate station* | *Marine station* | *Hydrometeorological station* | *Agrometeorological station* | *Urban station* | *Proxy station* |
|---|---|---|---|---|---|---|---|
| Air temperature | • | • | • | | • | • | |
| Soil temperature | | | | | • | | |
| Water temperature | | | • | • | | | |
| Precipitation | • | • | • | • | • | • | |
| Weather | | • | • | | • | • | |
| Clouds | | • | • | | • | • | |
| Pressure | | • | • | | • | • | |
| Visibility | | • | • | | • | • | |
| Humidity | | • | • | | • | • | |
| Wind | | • | • | | • | • | |
| Solar radiation | | • | | | • | • | |
| Sunshine | | • | | | • | • | |
| Salinity | | | • | | | | |
| Currents | | | • | | | | |
| Sea level | | | • | | | | |
| Waves | | | • | | | | |
| Air–sea momentum | | | • | | | | |
| Air–sea fluxes | | | • | | | | |
| Ice | | | • | • | | | |
| Dissolved oxygen | | | • | | | | |
| Nutrients | | | • | | | | |
| Bathymetry | | | • | | | | |
| Biomass | | | • | | | | |
| Streamflow | | | | • | | | |
| River stages | | | | • | | | |
| Sediment flow | | | | • | | | |
| Recharge | | | | • | | | |
| Evaporation | | | | • | • | • | |
| Soil moisture | | | | • | • | • | |
| Runoff | | | | • | • | | |
| Groundwater | | | | • | • | | |
| Plant development | | | | | | • | • |
| Pollen | | | | | | | • |

| *Element* | *Ordinary climate station* | *Principal climate station* | *Marine station* | *Hydrometeorological station* | *Agrometeorological station* | *Urban station* | *Proxy station* |
|---|---|---|---|---|---|---|---|
| Ice and sediment composition | | | | | | | • |
| Tree ring growth | | | | | | | • |
| Coral ring growth | | | | | | | • |
| Atmospheric chemicals | | | | | | • | |
| Particulate matter | | | | | | • | |

### 2.2.2 **Upper-air elements**

Upper-air observations are an integral component of the Global Observing System. The spectrum of climate activities that require upper-air observations includes monitoring and detecting climate variability and change, climate prediction on all timescales, climate modelling, studies of climate processes, data reanalysis activities, and satellite studies concerning calibration of satellite retrievals and radiative transfer.

The longest record of upper-air observations has been obtained from balloon-based instruments combined with ground tracking devices in a radiosonde network. These radiosonde measurements provide a database of atmospheric variables dating back to the 1930s, although coverage is generally poor before 1957. The radiosonde data record is characterized by many discontinuities and biases resulting from instrument and operational procedural changes and incomplete metadata. Satellite observations have been available since the 1970s, and some have been assembled and reprocessed to create continuous records. Just as the radiosonde record has deficiencies, however, the satellite data also suffer from, among other things, limited vertical resolution, orbit drift, satellite platform changes, instrument drift, complications with calibration procedures, and the introduction of biases through modifications of processing algorithms. Other upper-air measurements have come from moving platforms such as aircraft. Observations from some high-mountain locations have also been considered part of the upper-air measurement system.

The main observational requirements for monitoring long-term upper-air changes are:

(a) A long-term (multi-decadal), stable, temporally homogeneous record so that changes can confidently be identified as true atmospheric changes rather than changes in the observing system or as artefacts of homogenization methods;

(b) Good vertical resolution to describe the vertical structure of temperature, water vapour, and ozone changes, and of changes in the tropopause;

(c) Sufficient geographical coverage and resolution, so that reliable global and area trends can be determined;

(d) Observational precision that is finer than the expected atmospheric variations, so as to allow clear identification of both variability and long-term changes. This requirement is particularly important for water vapour observations in the upper troposphere and stratosphere.

The essential climate elements from upper-air observations are: temperature, water vapour, pressure, wind speed and direction, cloud properties, radiance and radiation (net, incoming and outgoing), as given in the *Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCCC* (GCOS-82, WMO/TD-No. 1143) and in the *Implementation*

*Plan for the Global Observing System for Climate in Support of the UNFCCC* (GCOS-92, WMO/TD-No. 1219). Because the chemical composition of the atmosphere is of great importance in climate prediction, climate change monitoring, ozone and other air-quality predictions, and in application areas such as the study and forecasting of animal, plant and human health and well-being, it is important to understand the vertical structure of the composition of the global atmosphere (see the *Technical Regulations*, Volume I, Part I, 6.3, the *Plan for the Global Climate Observing System (GCOS)* (GCOS-14, WMO/TD-No. 681) and the *GCOS/GTOS Plan for Terrestrial Climate-related Observations* (GCOS-32, WMO-No. 796)). Chemical composition elements requiring measurement both in the free atmosphere and near the ground include concentrations of ozone and other greenhouse gases, such as carbon dioxide and methane, atmospheric turbidity (aerosol optical depth), aerosol total load, reactive gas species, and radionuclides. Measurements of acid rain (or more generally precipitation and particulate chemistry) and ultraviolet radiation are also needed (see *The Changing Atmosphere: An Integrated Global Atmospheric Chemistry Observation Theme for the IGOS Partnership. GAW Report No. 159* (WMO/TD-No. 1235)) for details concerning the chemical composition of the atmosphere.

Upper-air measurements should capture the full range of climate regimes and surface types. Radiative transfer codes used to convert raw satellite radiances to geophysical parameters depend upon assumptions about the surface conditions. Therefore, different local environmental conditions should be represented, including both land and ocean areas.

### 2.2.3 Elements measured by remote sensing

Satellites and other remote-sensing systems such as weather radar provide an abundance of additional information, especially from otherwise data-sparse areas, but they are not yet capable of providing, with the required accuracy and homogeneity, measurements of many of the elements that are reported from land-based surface stations. The spatial coverage they offer makes them complementary to, but not a substitute for, surface networks. The elements that can be measured or estimated remotely are precipitation (with limited accuracy over small areas, ocean–atmosphere interfaces, highlands or steep orography); cloud amount; radiation fluxes; radiation budget and albedo; upper oceanic biomass, ocean surface topography and wave height; sea-ice cover; sea surface temperature; ocean surface wind vectors and wind speed; atmospheric temperature, humidity and wind profiles; chemical constituents of the atmosphere; snow cover; ice sheet and glacier extent; vegetation and land cover; and land surface topography.

Greater spatial and temporal coverage can be achieved with remote-sensing than with in situ observations. Remotely sensed data also supplement observations from other platforms and are especially useful when such observations are missing or corrupted. Although this is an advantage, there are problems in using remotely sensed data directly for climate applications. Most importantly, the short period of record means that remotely sensed data cannot be used to infer long-term climate variability and change. Also, remotely sensed data may not be directly comparable to in situ measurements. For example, satellite estimates of the Earth’s skin temperature are not the same as temperature measurements taken in a standard screen, and the relationship between radar measurements of reflectivity and precipitation amounts collected in rain gauges may be quite complex. It is possible with care, however, to construct homogeneous series that combine remotely sensed and in situ measurements.

## 2.3 INSTRUMENTATION

Climatological stations that are part of a national network should be equipped with standard approved instruments; the NMHS may supply the instruments. When equipment is supplied by other agencies or purchased by the observer, every effort should be made by a climate office to ensure compliance with national standards.

This section provides guidance on some basic surface instrumentation and on the selection of instruments. There are several other WMO publications that are necessary companions to this Guide; they should be readily available and consulted as needed. A thorough survey of instruments suitable for measuring climate and other elements at land and marine stations is provided in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). Details of instrumentation needed for the measurement of agrometeorological elements are given in the *Guide to Agricultural Meteorological Practices* (WMO-No. 134), and for hydrological purposes in the *Guide to Hydrological Practices* (WMO-No. 168).

When selecting instrumentation, including any associated data-processing and transmission systems, the ten climate monitoring principles (section 2.1) should be followed. Several concerns should be considered when complying with these principles:

– Reliability;

– Suitability for the operational environment at the station of use;

– Accuracy;

– Simplicity of design;

– Reasons for taking observations.

Reliability requires that an instrument functions within its design specifications at all times. Unreliable instrumentation leads to data gaps, biases and other inhomogeneities. Reliable instruments need to be robust enough to cope with the full range of weather and physical extremes expected at the site, and possibly the handling that is part of manual observations.

Instruments must be suited both to the climate in which they must function, and to other equipment with which they must operate. For example, an anemometer head at a cold location will need to withstand icing, while one in a desert area will need to be protected against dust ingression. Sensors for use in an automatic weather station need to provide output suitable for automatic processing. The standard mercury-in-glass thermometer used at a manual recording site, for example, will need to be substituted by a temperature-sensitive probe, such as a thermocouple, whose response can be converted into an electronic signal. Instruments must also be sited so that they can be accessed and maintained.

Ideally, all instruments should be chosen to provide the high level of accuracy and precision required for climatological purposes. It is also important that the instrument can continue to provide the required level of accuracy for a long period of time, as instrument "drift" can lead to serious inhomogeneities in a climate record; accuracy is of limited use without reliability.

The simpler the instrumentation, the easier it is to operate and to maintain, and the easier it is to monitor its performance. It is sometimes necessary to install redundant sensors (for example, triple thermistors at automated data-logger stations) to properly track performance and reliability over time. Complex systems can easily lead to data inhomogeneities, data loss, high maintenance cost and changing accuracy.

The purpose of the observations generally dictates requirements for measurements. Types of instrument, installation of sensors, and characteristics of the instruments should be considered to ensure that measurement requirements can be met. Details about these concerns, including instrument and measurement standards and recommended practices, can be found in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8).

### 2.3.1 **Basic surface equipment**

There may be a variety of options for obtaining climate observations from surface stations. These options include equipping a station with, for example, basic instruments, autographic or automated output available for unmanned periods, or totally automated sensors. When considering the options, it is important to compare costs of personnel, maintenance and replacement. Price negotiation is often possible with manufacturers on the basis of, for example, quantities purchased, among other things.

Whenever possible, trained local personnel, such as a caretaker, should examine an observing site on a regular basis to keep surface conditions (such as grass growth) in check, perform basic maintenance of instruments (such as simple cleaning), examine for damage, and detect breaches of security. These tasks should be performed at least weekly at accessible, manned land stations. Inspection of sites and instruments at remote locations should be made as often as possible. Personnel should also be available to provide rapid maintenance response when critical systems fail.

Autographic and data-logger equipment exists for the recording of many climatic elements, such as temperature, humidity, wind and rates of rainfall. Data need to be transferred from autographic records to tables or digital form. Observers should ensure that the equipment is operating properly and that information recorded on charts, for example, is clear and distinct. Observers should be responsible for regularly verifying and evaluating the recorded data (by checking against direct-reading equipment) and for making time marks at frequent, specified intervals. The recorded data can be effectively used to fill gaps and to complete the record when direct observations are missed because of illness and other causes of absence from the observing station. The *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8) gives specific guidance on the maintenance and operation of recording instruments, drums and clocks.

Data from automatic weather stations (AWSs), at which instruments record and transmit observations automatically, are usually restricted to those readily obtained in digital form, although the range of sensors is wide and continues to evolve. Such stations have been used to supplement manned stations and to increase network densities, reporting frequencies and the quantities of elements observed, especially in remote and largely unpopulated areas where human access is difficult. Some of the sensitivity and accuracy requirements of these automated stations are given in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8); others are being developed, especially for studies of climate variability.

In many countries, AWSs have lowered operational costs. National Meteorological and Hydrological Services choosing between manned and AWS observation programmes need to consider a number of issues. Notwithstanding the considerable potential for AWSs to provide high-frequency data, as well as additional data from remote locations, there are several significant costs associated with operating an AWS, including labour costs for maintenance and ensuring AWS reliability, labour availability, accessibility for installation and maintenance, availability of suitable power sources, security of the site, and communications infrastructure. These aspects must be carefully weighed against the significant benefits, such as a denser or more extensive network. Automatic weather stations can be powerful alternatives to manned observational programmes and sometimes are the only option, but they require a strong organizational commitment to manage them.

Marine instrumentation includes drifting and moored data buoys, ice floats and subsurface floats (such as Argo float). Although the data are collected remotely, the instruments are generally performing in situ measurements. They are a cost-effective means for obtaining meteorological and oceanographic data from remote ocean areas. As such, they form an essential component of marine observing systems and meteorological and oceanographic operational and research programmes. For example, the Tropical Atmosphere Ocean array of moorings has enabled timely collection of high-quality oceanographic and surface meteorological data across the equatorial Pacific Ocean for the monitoring, forecasting and understanding of climate swings associated with El Niño and La Niña.

### 2.3.2 **Upper-air instruments**

Historically, most climatological data for the upper air have been derived from measurements made for synoptic forecasting by balloon-borne radiosondes. A variety of techniques and instruments are used for the measurement of pressure, temperature, humidity and wind, and for processing instrumental output into meteorological quantities. It is important for each NMHS to issue suitable instruction manuals to each upper-air station for the proper use of equipment and interpretation of data. The *Manual on the Global Observing System* (WMO-No. 544) requires that prompt reports be made to the WMO Secretariat of changes in radiosonde type or changes in wind systems in operational use at a station.

There are several issues concerning the quality of radiosonde measurements for climate monitoring and climate change detection purposes. Radiation errors cause uncertainties in temperature. Standard radiosondes are not capable of measuring water vapour at low temperatures with sufficient accuracy. Sensor types, especially humidity sensors, have changed over time. The spatial coverage of radiosonde observations is not uniform; most stations are located on the land surface of the northern hemisphere, while the southern hemisphere and ocean networks are much less dense. The historical record of radiosonde observations has innumerable problems relating to a lack of inter-comparisons among types of radiosondes, and sensor and exposure differences; metadata concerning instrumentation, data-reduction and data-processing procedures are crucial to utilizing radiosonde data in climate applications. New reference radiosondes are being developed to mitigate the deficiencies of the current standard radiosondes. A limited network of these will be used to calibrate and validate various satellite observations of both temperature and water vapour.

An upper-air observing system may change over time with technological advances. Hence, a key requirement of the network is sufficient overlap of systems to maintain continuity and allow full comparison of the accuracy and precision of the old and new systems. Measurement systems should be calibrated regularly at the site. It is imperative that instrument replacement strategies take into account changes in other networks, such as the use of satellites. The climate monitoring principles (see 2.1) should guide the development and operation of an upper-air observing system.

### 2.3.3 **Surface-based remote sensing**

Remote sensing can use either active or passive sensors. Active sensor systems emit some form of radiation, which is scattered by various targets; the sensors detect the backscatter. Passive sensors measure radiation being emitted (or modified) by the environment.

The most common surface-based active remote-sensing technique is weather radar. A short pulse of high-power microwave energy is focused by an antenna system into a narrow beam. This beam is scattered back by the target precipitation, with the backscattered radiation received, generally, by the same antenna system. The location of the precipitation can be determined from the azimuth and elevation of the antenna and the time between transmitting and receiving the reflected energy. The power of the received radiation depends on the nature of the precipitation, and the signal can be processed to estimate its intensity. Atmospheric and environmental conditions can adversely affect radar data, and caution should be exercised when interpreting the information. Some of these effects include returns from mountains, buildings and other non-meteorological targets; attenuation of the radar signal when weather echoes are viewed through intervening areas of intense precipitation; temperature inversions in the lower layers of the atmosphere, which bend the radar beam in such a way that ground clutter is observed where normally not expected; and the bright band, which is a layer of enhanced reflectivity caused by the melting of ice particles as they fall through the freezing level in the atmosphere, which can result in overestimation of rainfall. Use of radar data in climate studies has been limited by access and processing capabilities, uncertainties in calibration and calibration changes, and the complex relationship between reflectivity and precipitation.

Wind profilers use radar to construct vertical profiles of horizontal wind speed and direction from near the surface to the tropopause. Fluctuations of atmospheric density are caused by turbulent mixing of air with different temperatures and moisture content. Fluctuations in the resulting index of refraction are used as a tracer of the mean wind. Although they work best in clear air, wind profilers are capable of operating in the presence of clouds and moderate precipitation. When equipped with a radio-acoustic sounding system, profilers can also measure and construct vertical temperature profiles. The speed of sound in the atmosphere is affected by temperature. Acoustic energy is tracked through the atmosphere, and the temperature profile is estimated from the speed of the sound wave propagation.

Lightning detection is the most common passive surface-based remote-sensing. Lightning sensors scan a range of electromagnetic frequencies to detect electrical discharges inside clouds, between clouds, or between clouds and the ground. Characteristics of the received radiation (such as the amplitude, time of arrival, source direction, sign and other wave form characteristics) are measured, and from them characteristics of the lightning flash are inferred. One sensor cannot accurately locate lightning events; data from several sensors are concentrated in a central location in a central lightning processor. The processor computes and combines data from multiple sensors to calculate the location and characteristics of the observed lightning flashes. The accuracy and efficiency of a lightning-detector network drops progressively on its outer boundaries. The detection wave will propagate without too much attenuation over distance, depending on the frequency band used, but if a lightning flash is too far away from the network (this distance varies with the stroke amplitude and the network configuration), the stroke may no longer be detected.

### 2.3.4 **Aircraft-based and space-based remote sensing**

Many long-distance aircraft are fitted with automatic recording systems that report temperature and wind, and in some cases humidity, regularly while *en route*. Some aircraft record and report frequent observations during take-off and descent to significantly augment the standard radiosonde data, at least throughout the troposphere. Such data are assimilated into operational meteorological analysis systems and, through programs of reanalysis, ultimately contribute substantially to the broader climate record.

Aircraft meteorological data-relay systems operate on aircraft that are equipped with navigation and other sensing systems. There are sensors for measuring air speed, air temperature and air pressure. Other data relating to aircraft position, acceleration and orientation are obtained from the aircraft navigation system. The aircraft also carry airborne computers for flight management, and navigation systems by which navigation and meteorological data are computed continuously and are made available to the aircrew. The data are automatically fed to the aircraft communication system for transmission to the ground, or alternatively, a dedicated processing package can be used on the aircraft to access raw data from the aircraft systems and derive the meteorological variables independently. Normally, messages transmitted to ground stations contain horizontal wind speed and direction, air temperature, altitude (related to a reference pressure level), a measure of turbulence, time of observation, phase of flight, and the aircraft position. The data are used by aviation controllers to ensure flight safety and by weather forecasters.

There are potentially a large number of error sources contributing to aircraft measurement uncertainty. An uncertainty of about 5 to 10 per cent in the calculation process can be expected. A further complication arises over the choices of sampling interval and averaging time. Examination of typical time series of vertical acceleration data often indicates a high variability of statistical properties over short distances. Variation of air speed for a single aircraft and between different aircraft types alters the sampling distances and changes the wavelengths filtered. While not as precise and accurate as most ground observing systems, aircraft data can provide useful supplemental information to meteorological databases.

Satellite data add valuable information to climate databases due to their wide geographical coverage, especially over areas with sparse or completely missing in situ data. Satellites are very useful for monitoring phenomena such as polar sea-ice extent, snow cover, glacial activity, sea-level changes, vegetation cover and moisture content, and tropical cyclone activity. They also help improve synoptic analyses, an important component of synoptic climatology.

Sensing techniques make use of the emission, absorption and scattering properties of the atmosphere and the Earth's surface. The physical equations for radiative transfer provide information about the radiative properties of the atmosphere and the Earth's surface and, through inversion of the radiative transfer equation, geophysical properties such as temperature and moisture profiles, surface skin temperature, and cloud properties.

The figures and specifications of satellite platforms and sensors are in the *Statement of Guidance Regarding How Well Satellite Capabilities Meet WMO User Requirements in Several Application Areas* (SAT-22, WMO/TD-No. 992). The elements, accuracy and spatial and temporal resolution of data measured by satellites can be found in the following publications: *Report of the Capacity-building Training Workshop on Reducing the Impacts of Climate Extremes on Health* (WCASP-No. 59, WMO/TD-No. 1162), Chapter 1; *Guideline for the Generation of Satellite-based Datasets and Products Meeting GCOS Requirements* (GCOS-128, WMO/TD-No. 1488), section 4; and the above-mentioned WMO/TD-No. 992, section 3.1 and Annex A.

The histories and future plans of satellite platforms and sensors are in *Systematic Observation Requirements for Satellite-based Products for Climate* (GCOS-107, WMO/TD-No. 1338). The technology of remote-sensing is progressing rapidly and the operational plans of platforms and sensors may be changed occasionally. Therefore, the latest documents should be referred to in using remote-sensing data. Reports published by the Committee on Earth Observation Satellites, available on the Internet, are helpful in seeking the latest information about satellites.

As in the case of surface-based remote-sensing, satellite and other airborne sensors can be classified into two groups: passive and active. Passive sensors include imagers, radiometers and sounders. They measure radiation emitted by the atmosphere or the Earth's surface. Their measurements are converted into geophysical information such as vertical profiles of water vapour, temperature and ozone; cloud information; surface ocean and land temperatures; and ocean and land colour. The wavelength at which a sensor operates influences the resulting information, with different wavelengths having different advantages and disadvantages.

Active sensors include radar, scatterometers and lidar. They measure the backscattered signal from an observing target when it is illuminated by a radiation source emitted from the platform. Their advantage is that the accurate range of an observing target can be obtained by measuring a time lag between an emission and its return, while the use of a tightly focused and directional beam can provide positional information. Backscattered signals can be converted into wind speed and direction, ocean dynamic height and wave spectrum, ocean wind stress curl and geostrophic flow, cloud properties, precipitation intensity, and inventories of glacial extent.

Sometimes, information can be derived from satellite data that was not originally intended for climatological purposes. For example, the Global Positioning System uses a network of dozens of satellites to assist in navigation. But by measuring the propagation delay in Global Positioning System signals, it is possible to estimate atmospheric water vapour content.

Two complementary orbits have been used for operational environmental satellites: geostationary and polar orbits. A geostationary satellite, placed at about 36 000 km above the Equator, will orbit the Earth once every 24 hours. The satellite, therefore, remains stationary relative to the Earth and can thus provide a constant monitoring capability and the ability to track atmospheric features and infer winds. Polar-orbiting satellites are typically about 800 km above the surface, moving almost north–south relative to the Earth. Most of the globe is observed by the suite of instruments on the operational polar-orbiting satellites twice per day, about 12 hours apart. The inconvenience of only two passes each day is balanced by the higher spatial resolution and greater range of instruments carried, and the ability to see high latitudes that are poorly captured from the geostationary orbit.

The climate community must recognize the need to provide scientific data stewardship of both the "raw" remotely-sensed measurements and the data processed for climate purposes. In addition to the ten principles listed in 2.1, satellite systems should also adhere to the following principles:

1. Constant sampling within the diurnal cycle (minimizing the effects of orbital decay and orbit drift) should be maintained;

2. Overlapping observations should be ensured for a period sufficient to determine inter-satellite biases;

3. Continuity of satellite measurements (elimination of gaps in the long-term record) through appropriate launch and orbital strategies should be ensured;

4. Rigorous pre-launch instrument characterization and calibration, including radiance confirmation against an international radiance scale provided by a national metrology institute, should be ensured;

5. On-board calibration suitable for climate system observations should be ensured and associated instrument characteristics should be monitored;

6. Operational generation of priority climate products should be sustained, and peer-reviewed new products should be introduced as appropriate;

7. Data systems needed to facilitate users' access to climate products, metadata and raw data, including key data for delayed-mode analysis, should be established and maintained;

8. Use of functioning baseline instruments that meet the calibration and stability requirements stated above should be maintained for as long as possible, even when these exist on decommissioned satellites;

9. Complementary in situ baseline observations for satellite measurements should be maintained through appropriate activities and cooperation;

10. Random errors and time-dependent biases in satellite observations and derived products should be identified.

The *Manual on the Global Observing System* (WMO-No.544) provides guidance on various components of the observing systems.

### 2.3.5 **Calibration of instruments**

It is of paramount importance, for determining the spatial and temporal variations of climate, that the relative accuracy of individual sensors in use in a network at one time be measured and periodically checked and, similarly, that the performance of replacement sensors and systems can be related to that of the instruments replaced. The *Manual on the Global Observing System* (WMO-No. 544) states that all stations shall be equipped with properly calibrated instruments. Details on calibration techniques can be found in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). For climatology, it is not generally sufficient to rely upon manufacturers' calibrations and it is wrong to assume that a calibration will not drift or otherwise change with time.

Comparisons of instrumental or system measurements should be made with portable standard instruments when replacement instruments are issued to a station and at each regular inspection of the station (see 2.6.6). Travelling standards should be checked against national reference standards before and after each period of travel, and they should be robust in transport and withstand calibration changes. Records of instrument changes and calibration drifts must be kept and made available as metadata, as they are essential to the assessment of true climate variations (see 2.6.9).

During inspections of remotely sited AWSs, observations should be taken using the travelling standards for later comparison with the recorded AWS output received at the data reception point. Some NMHSs have automated fault or instrumental drift detection procedures in place, which compare individual measurements with those from a network and with values analysed from numerically fitted fields. These automated procedures are useful for detecting not only drift, but also anomalous step changes.

Some NMHSs operate their own calibration facilities or use accredited calibration companies. Regional calibration facilities within WMO are responsible for keeping and calibrating standards, certifying an instrument's conformity to standards, organizing instrument evaluations, and providing advice about instrumental performance.

The GCOS *Plan for Space-based Observations* details calibration, inter-calibrational overlapping records and metadata requirements for space-based remote sensors. The plan of the Global Space-based Inter-calibration System is to compare the radiances simultaneously measured by satellite pairs at the crossing points of their ground track, in particular where a polar orbiter and a geostationary satellite cross paths. This inter-calibration will give a globally consistent calibration on an operational basis.

Weather radar calibration requires the measurement of system characteristics such as transmitted frequency and power, antenna gain, beam widths, receiver output and filter losses. Performance monitoring ensures that other system characteristics, such as antenna orientation, side lobes, pulse duration and pulse shape, beam patterns and receiver noise levels, are within acceptable limits.

Drifts of lightning detection sensors or central lightning processor parameters are problems that should be detected through regular data analyses (for example, cross-checking of sensor behaviour and analysis of stroke parameters). Comparison should also be made with other observations of lightning activity, such as manual observations of "thunder heard" or "lightning seen", or observations of cumulonimbus clouds. As with weather radars, monitoring and calibration of the characteristics of the system should be a routine process.

## 2.4 SITING OF CLIMATOLOGICAL STATIONS

The precise exposure requirements for specific instruments used at climatological stations, aimed at optimizing the accuracy of the instrumental measurements, are discussed in the *Manual on the Global Observing System* (WMO-No.544), Part III, the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No.8), and in *Representativeness, Data Gaps and Uncertainties in Climate Observations* (WMO/TD-No. 977). These publications are a necessary companion to this Guide.

The representativeness and homogeneity of climatological records are closely related to the location of the observing site. A station sited on or near a steep slope, ridge, cliff, hollow, building, wall or other obstruction is likely to provide data that are more representative of the site alone not of a wider area. A station that is or will be affected by the growth of vegetation, including even limited tree growth near the sensor, growth of tall crops or woodland nearby, erection of buildings on adjacent land, or increases (or decreases) in road or air traffic (including those due to changes in the use of runways or taxiways) will provide neither broadly representative nor homogeneous data.

A climatological observing station should be sited at a location that permits the correct exposure of the instrumentation and allows for the widest possible view of the sky and surrounding country, if visual data are required. Ordinary and principal climatological stations should be sited on a level piece of ground covered with short grass; the site should be well away from trees, buildings, walls and steep slopes and should not be in a hollow. A plot size of about 9 metres by 6 metres is sufficient for outdoor temperature and humidity-sensing instruments, and an area of 2 metres by 2 metres of bare ground within the plot is ideal for observations of the state of the ground and soil temperature measurements. A slightly larger plot (10 metres by 7 metres) is preferable if the site is to enclose a rain gauge in addition to the other sensors.

A rule used by many NMHSs is that the distance of any obstruction, including fencing, from the rain gauge must be more than twice, and preferably four times, the height of the object above the gauge. In general terms, anemometers should be placed at a distance from any obstruction of at least 10, and preferably 20, times the height of the obstruction. The different exposure requirements of various instruments may give rise to a split site, where some elements are observed from one point while others are observed nearby, with data from all the elements combined under the one site identifier.

Prevention of unauthorized entry is a very important consideration and may require enclosure by a fence. It is important that such security measures do not themselves compromise the site exposure. Automatic stations will normally need a high level of security to protect against animal and unauthorized human entry; they also require the availability of suitable and robust power supplies, and may possibly need additional protection against floods, leaf debris and blowing sand.

Ordinary and principal climatological stations should be located at such sites and should be subject to such administrative conditions that will allow the continued operation of the station, with the exposure remaining unchanged, for a decade or more. For stations used or established to determine long-term climate change, such as reference climatological stations and other baseline stations in the GCOS network, constancy of exposure and operation is required over many decades.

Observing sites and instruments should be properly maintained so that the quality of observations does not deteriorate significantly between station inspections. Routine, preventive maintenance schedules include regular "housekeeping" at observing sites (for example, grass cutting and cleaning of exposed instrument surfaces, including thermometer screens) and manufacturers' recommended checks on instruments. Routine quality control checks carried out at the station or at a central point should be designed to detect equipment faults at the earliest possible stage. Depending on the nature of the fault and the type of station, the equipment should be replaced or repaired according to agreed priorities and time intervals. It is especially important that a log be kept of instrument faults and remedial action taken where data are used for climatological purposes. This log will be the principal basis for the site's metadata and hence will become an integral part of the climate record. Detailed information on site maintenance can be found in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8, 1.2.5.2).

Additional constraints on siting apply to GAW stations established to provide data on atmospheric chemical composition, as discussed in the *Technical Regulations* (WMO-No. 49), Volume I, Part I, 6.1). These constraints include the need for no significant changes in land-use practices within 50 kilometres of the site, and freedom from the effects of local and area pollution from, for example, major population centres, industrial and extensive farming activities, highways, volcanic activity and forest fires. Both global and regional GAW stations should be within 70 kilometres of an upper-air synoptic station.

The nature of urban environments makes it impossible to conform to the standard guidance for site selection and exposure of instrumentation required for establishing a homogeneous record that can be used to describe the larger-scale climate. Nonetheless, urban sites do have value in their own right for monitoring real changes in local climate that might be significant for a wide range of applications. Guidelines for the selection of urban sites, installation of equipment and interpretation of observations are given in *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites* (WMO/TD-No. 1250). Fundamental to the guidance is the need to clearly understand the purpose of making the observations and to obtain measurements that are representative of the urban environment. In many urban situations, it will be possible to conform to standard practices, but flexibility in siting urban stations and in instrumentation may be necessary. These characteristics further heighten the importance of maintaining metadata that accurately describe the setting of the station and instrumentation.

## 2.5 THE DESIGN OF CLIMATOLOGICAL NETWORKS

A network of stations is several stations of the same type (such as a set of precipitation stations, radiation measuring stations or climatological stations), which are administered as a group. Each network should be optimized to provide the data and perform as required at an acceptable cost.

Most optimizing methods rely on data from a pre-existing network, available over a long enough period to correctly document the properties of the meteorological fields. They are based on both temporal and spatial statistical analyses of time series. It is difficult to assess a priori how long the data series must be because the number of years necessary to capture variability and change characteristics may vary with the climatic element. It has been common practice to assume that at least ten years of daily observations are necessary to produce the relevant base statistical parameters for most elements, and at least thirty years for precipitation. Observed global and regional climatic trends and variability in many areas of the globe over the past century suggest, however, that such short periods of record may not be particularly representative of similar periods to follow.

The identification of redundant stations allows network managers to explore options for optimizing the network, for example, by eliminating the redundant stations to reduce costs or by using the resources to establish stations at locations where observations are needed for a more effective realization of the network objectives. Network managers should take advantage of the relatively high spatial coherence that exists for some meteorological fields, such as temperature. Techniques used to evaluate the level of redundancy of information include the use of the spatial variance–covariance matrix of the available stations, multiple linear regression, canonical analysis and observation system simulation experiments (see Chapter 5).

The density and distribution of climatological stations to be established in a land network within a given area depend on the meteorological elements to be observed, the topography and land use in the area, and the requirements for information about the specific climatic elements concerned. The rate of variation of climatic elements across an area will differ from element to element. A sparse network is sufficient for the study of surface pressure, a fairly dense network for the study of maximum and minimum temperature, and very dense networks for examining the climatology of precipitation, wind, frost and fog, especially in regions of significant topography.

Stations should be located in such a way as to give representative climatic characteristics that are consistent with all types of terrain, such as plains, mountains, plateaus, coasts and islands, and surface cover such as forests, urban areas, farming areas and deserts within the area concerned. Station density should be determined by the purposes of the observations and the uses of the data. For data used in sectoral applications within an area, there may be a need for a greater density of stations where human activities or health are sensitive to climate, and a lesser density in locations with fewer people. When planning a land network, compromises often have to be made between the ideal density of stations and the resources available to install, operate and administer the stations.

The distribution of stations in the Regional Basic Synoptic Network (RBSN), from which monthly surface climatological data are collected, should be dense enough to provide an accurate description of the atmosphere for those who use observations.

The network of climatological stations should give a satisfactory representation of the climate characteristics of all types of terrain in the territory of the Member concerned (for example, plains, mountainous regions, plateaus, coasts and islands). Regional Basic Synoptic Networks of both surface and upper-air stations, and Regional Basic Climatological Networks (RBCNs) of climatological stations shall be established to meet the requirements laid down by the regional associations (see *Manual on the Global Observing System* (WMO-No. 544), Part III, 2.1.3.2).

Each Member should establish and maintain at least one reference climatological station for determining climate trends. Such stations need to provide more than 30 years of homogeneous records and should be situated where anthropogenic environmental changes have been and are expected to remain at a minimum. Information on agrometeorological and hydrometeorological networks and sites can be found in the *Guide to Agricultural Meteorological Practices* (WMO-No. 134) and the *Guide to Hydrological Practices* (WMO-No. 168), respectively, and additional guidance is provided in the *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160).

A country's environmental information activities are often conducted by many parties whose contributions are complementary and at times overlapping. A country benefits from environmental information collected and disseminated by both governmental agencies and non-governmental entities (including private companies, utilities and universities). Formal partnerships between the NMHS and these other parties are highly desirable for optimizing resources. Because data and information obtained from non-NMHS sources are not usually under the control of the NMHS, metadata are critical for the most effective use of the information. As for stations maintained by the NMHS, metadata on instrumentation, siting, processing procedures, methodologies and anything else that would enhance the use of the information should be obtained and documented. The metadata should also be maintained and be accessible. To promote the open and unrestricted exchange of environmental information, including weather observations, it is highly desirable that the NMHS be granted full use of all the climate data and information obtained from partnerships, without restriction, as if they were its own data. An appropriate contract or memorandum of understanding between the NMHS and other organizations may need to be drafted and signed at the senior management level.

In addition to data from standard and private networks of climatological stations, there are sometimes observational data from networks of temporary stations established in connection with research and study programmes, as well as measurements made in mobile transects and profiles. The NMHS should endeavour to obtain these data and associated metadata. Although the data may not be ideal for typical archiving, they will often prove to be quite valuable as supplementary information, for example, for investigations of specific extreme events. When these observations are collected from data-poor areas, they are highly valuable.

## 2.6 STATION AND NETWORK OPERATIONS

Guidance material in this section concerns mainly observations at ordinary climatological stations (at which observations are usually made twice a day, but in some cases only once a day, and include readings of extreme temperature and precipitation). Guidance is also given regarding precipitation stations (stations at which one or more observations of precipitation only are made each day). Regulatory and guidance material for principal climatological stations (which usually also function as synoptic observing stations) and other types of climatological stations can be found in the *Manual on the Global Observing System* (WMO-No. 544).

### 2.6.1 Times of observations

Observations at ordinary climatological and precipitation stations should be made at least once (and preferably twice) each day at fixed hours that remain unchanged throughout the year. At principal climatological stations, observations must be made at least three times daily in addition to an hourly tabulation from autographic records, but non-autographic observations are usually taken hourly. From a practical viewpoint, times of observation should fit the observer’s working day, usually one morning observation and one afternoon or evening observation. If daylight saving time is used for part of the year, the observations should continue to be made according to the fixed local time; the dates when daylight saving time commences and ends must be recorded. If at all possible, the times of observation should coincide with either the main or intermediate standard times for synoptic observations (0000, 0300, 0600 Coordinated Universal Time (UTC), and so on). If conditions dictate that only one observation a day is possible, this observation should be taken between 0700 and 0900 local standard time.

In selecting the schedule for climatological observations, times at or near the normal occurrence of daily minimum and maximum temperatures should be avoided. Precipitation amounts and maximum temperatures noted at an early morning observation should be credited to the previous calendar day, while maximum temperatures recorded at an afternoon or evening observation should be credited to the day on which they are observed.

Times of observation often vary among networks. Summary observations such as temperature extremes or total precipitation made for one 24-hour period (such as from 0800 on one day to 0800 on the next day) are not equivalent to those made for a different 24-hour period (such as from 0000 to 2400).

If changes are made to the times of observations across a network, simultaneous observations should be carried out at a basic network of representative stations for a period covering the major climatic seasons in the area at the old and new times of observation. These simultaneous observations should be evaluated to determine if any biases result from the changed observation times. The station identifiers for the old and new times of observations must be unique for reporting and archiving.

### 2.6.2 Logging and reporting of observations

Immediately after taking an observation at a manual station, the observer must enter the data into a logbook, journal or register that is kept at the station for this purpose. Alternatively, the observation may be entered or transcribed immediately into a computer or transmission terminal and a database. Legislation or legal entities (such as courts of law) in some countries may require that a paper record or a printout of the original entry be retained for use as evidence in legal cases, or there may be difficulties associated with the acceptance of database-generated information. The observer must ensure that a complete and accurate record has been made of the observation. At a specified frequency (ranging from immediately to once a month), depending on the requirements of the NMHS, data must be transferred from the station record (including a computer database) to a specific report form for transmittal, either by mail or electronically, to a central office.

Climatological station personnel must ensure that there is a correct copy of the pertinent information in the report form. In the case of paper records, the need for good, clear handwriting and "clean" journals and report forms should be emphasized. It is quite common for more information, perhaps pertaining to unusual weather phenomena and occurrences, to be entered in the local record than is required by the central office. The on-station record must be retained and readily accessible so that the station personnel can respond to any inquiries made by the central office regarding possible errors or omissions in the report form. Some services request observers to send logbooks to the national climate centre for permanent archiving.

Some national climate centres require that station personnel calculate and insert monthly totals and means of precipitation and temperature so that the data may be more easily checked at the section or central office. In addition, either the climate centre or observer should encode data for the CLIMAT messages, as described in the *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No.1188), if appropriate. Software to encode the data has been developed by WMO. The observer should note in the station logbook and on the report forms the nature and times of occurrence of any damage to or failure of instruments, maintenance activities, and any change in equipment or exposure of the station, since such events might significantly affect the observed data and thus the climatological record. Where appropriate, instructions should be provided for transmitting observations electronically. If mail is the method of transmission, instructions for mailing should be provided to the station, as well as pre-addressed, stamped envelopes for sending the report forms to the central climate office.

### 2.6.3 On-site quality control

General guidance on on-site quality control of observations and reports is given in the *Manual on the Global Observing System* (WMO-No. 544), Part III, 2.7.3, and detailed guidance is given in the *Guide to the Global Observing System* (WMO-No. 488), Part III, 3.1.3.14, and Part VI. The procedures described below should be followed when there is an observer or other competent personnel on site.

Checks should be made for gross errors, against existing extremes, for internal consistency in a sequence of observations, for consistency in the sequence of dates and times of observation, for consistency with other elements and calculations, and for the accuracy of copies and of encoded reports. These checks can be done either manually or by using automated procedures. If there are errors, remedial action such as correcting the original data and the report should be taken before transmission. Errors detected after transmission should also be corrected and the corrected report should be retransmitted. Checks should also be made, and any necessary amendments recorded and corrections transmitted, if a query about data quality is received from an outside source. Records of an original observation containing an error should include a notation or flag indicating that the original value is erroneous or suspect. On-site quality control must also include maintenance of the standard exposure of the sensors, of the site, and of the proper procedures for reading the instrumentation and checking autographic charts.

Any patterns of measurement error should be analysed, for example, to see if they relate to instrument drift or malfunction, and summaries of data or report deficiencies should be prepared monthly or annually.

### 2.6.4 Overall responsibility of observers

In general, the NMHS of each Member will specify the responsibilities of observers. The responsibilities should include the competent execution of the following:

(a) Making climatological observations to the required accuracy with the aid of appropriate instruments;

(b) Performing appropriate quality checks;

(c) Coding and dispatching observations in the absence of automatic coding and communication systems;

(d) Maintaining in situ recording devices and electronic data loggers, including the changing of charts when provided;

(e) Making or collating weekly or monthly records of climatological data, especially when automatic systems are unavailable or inadequate;

(f) Providing supplementary or backup observations when automatic equipment does not observe all required elements, or when the equipment is out of service.

### 2.6.5 Observer training

Observers should be trained or certified by an appropriate meteorological service to establish their competence to make observations to the required standards. They should have the ability to interpret instructions for the use of instrumental and manual techniques that apply to their own particular observing systems. Guidance on the instrumental training requirements for observers is given in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8).

Often, observers are either volunteers or part-time employees, or take observations as part of their other duties. They may have little or no training in climatology or in taking scientific observations, and thus will depend on a good set of instructions. Instructional booklets for ordinary climatological and precipitation station observers should be carefully prepared and made available to observers at all stations. The instructions should be unambiguous and should simply outline the tasks involved; they should provide only the information that the observer actually needs in order to perform the tasks satisfactorily. Illustrations, graphs and examples could be used to stimulate the interest of the observer and facilitate the understanding of the tasks to be undertaken every day. Sample copies of correctly completed pages of a logbook or journal and of a report form should be included in the instruction material available to an observer. Ideally, a climate centre representative should visit the site, install the station and instruct the observer.

An observer must gain familiarity with the instruments, and should be aware in particular of the sources of possible error in reading them. The instructions should include a descriptive text with simple illustrations showing the functioning of each instrument. Detailed instructions regarding methods to be used for day-to-day care, simple instrument maintenance and calibration checks should be given. If correction or calibration tables are necessary for particular observing and recording tasks, the observer should be made thoroughly familiar with their use. Instructions should also cover the operation of computer terminals used for data entry and transmission.

Instructions must cover visual as well as instrumental observations. Visual observations are particularly prone to subjective error and their accuracy depends on the skill and experience acquired by the observer. Since it is very difficult to check the accuracy or validity of an individual visual observation, as much guidance as possible should be given so that correct observations can be made.

To complement the instruction material, personnel responsible for station management in the climatological service should contact observing stations regarding any recurring observing errors or misinterpretation of instructions. Regular inspection visits provide the opportunity to address siting or instrument problems and to further the training of the observer.

Some climate centres arrange special training courses for groups of volunteer observers. Such courses are especially useful in creating a uniform high standard of observations, as a result of the training given and the availability of time to address a wider range of problems than may be raised by a single observer at an on-site visit.

### 2.6.6 Station inspections

Principal climatological stations should be inspected once a year. Ordinary climatological stations and precipitation stations should be inspected at least once every three years, or more frequently if necessary, to ensure the maintenance and correct functioning of the instruments and thus a high standard of observations. Automated stations should be inspected at least every six months. Special arrangements for the inspection of ship-based instruments are described in the *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8).

Before each inspection, the inspector should determine to the fullest extent possible the quality of information and data received from each station on the itinerary. At each inspection, it should be confirmed that:

(a) The observer's training is up to date;

(b) The observer remains competent;

(c) The siting and exposure of each instrument are known, recorded and still the best obtainable;

(d) The instruments are of an approved pattern, in good order and verified against relevant standards;

(e) There is uniformity in the methods of observation and procedures for calculating derived quantities from the observations;

(f) The station logbook is well maintained;

(g) The required report forms are sent punctually and regularly to the climate centre.

Inspection reports should include sketches, photographs or diagrams of the actual observing site, indicating physical objects that might influence the observed values of the climatic elements. The reports must also list any changes in instruments and any differences in readings between instruments and travelling standards, changes in exposure and site characteristics from the previous visit, and dates of appropriate comparisons and changes. Inspectors must also be prepared to advise observers on any problems arising in the transmission of data, including automated data-entry and transmission systems. Inspection reports are an important source of metadata for use in determining the homogeneity of a climate record and should be retained indefinitely, or the information therein should be transferred to a computerized database (see 3.1).

### 2.6.7 **Preserving data homogeneity**

Unlike observations taken solely to support the preparation of forecasts and warnings, a continuous, uninterrupted climate record is the basis for many important studies involving a diverse array of climatological communities. Homogeneous climate datasets are of the utmost importance for meeting the needs of climate research, applications and user services.

Changes to a site, or its relocation, are major causes of inhomogeneities. The ten principles of climate monitoring (see 2.1) should be followed when relocation of a climatological station is necessary, when one station is to be replaced by another nearby, or when instrument systems change. Where feasible and practical, both the old and new observing stations and instrumentation should be operated for an overlapping period of at least one year, and preferably two or more years, to determine the effects of changed instruments or sites on the climatological data. The old and new sites should have unique station identifiers for both reporting and archiving. Specific guidance is given in the *Guidelines for Managing Changes in Climate Observation Programmes* (WMO/TD-No. 1378).

### 2.6.8 **Report monitoring at collection centres**

Data collection or archiving centres need to check the availability and quality of information at the time when it is due from observers, and they should have additional responsibilities concerning data from automated measuring or transmission systems. Since such centres normally process large volumes of information, the use of computerized checking systems would save much effort.

The first task is to check that the expected observations have arrived and that they have been submitted at the correct time. If the expected observations are not available, the observer should be contacted to determine the reason. In the case of automated systems, “caretakers” must provide information on visible signs of failure as soon as possible to the authority responsible for maintenance of the observing and transmission systems.

Checks on the quality of data received from manned or automated sites should include those described in 2.6.3. Other checks are useful and can be readily made in computerized monitoring. They include checks against data from neighbouring stations, a variety of statistical checks, checks against preset limits, temporal consistency and inter-element consistency. Chapters 4 and 5 describe some of the techniques for checking data.

Monitoring shortly after observations are taken, either on site or remotely, is of limited value unless action is initiated to quickly remedy problems. Information must be fed back to the observers, caretakers, inspectors, and instrument or system maintainers or manufacturers, and information on the actions taken must then be fed back to the monitoring centre. Copies of all reports must be kept.

### 2.6.9 **Station documentation and metadata**

Observations without metadata are of very limited use; it is only when accompanied by adequate metadata (data describing the data) that the full potential of the observations can be utilized. These metadata are essential and should be kept current and be easily obtainable in the form of station catalogues, data inventories and climate data files. The World Meteorological Organization has developed metadata standards based on those of the International Organization for Standardization (ISO), especially the ISO 19100 series (see *WIGOS Metadata Standard* (WMO-No. 1192)). These standards are part of the WMO Integrated Global Observing

System (WIGOS), which provides the framework for all WMO observing systems and for WMO contributions to co-sponsored observing systems in support of all WMO Programmes and activities (see the *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160)).

WIGOS observations consist of an exceedingly wide range of data, from manual observations to complex combinations of satellite hyper-spectral frequency bands, measured in situ or remotely, from a single dimension to multiple dimensions, and those involving processing. A comprehensive metadata standard covering all types of observation is by nature complex to define. A user should be able to use the metadata to identify the conditions under which the observation (or measurement) was made, and any aspects that may affect its use or understanding.

Basic station metadata should include the station name and index number (or numbers); geographical coordinates; the elevation above mean sea level; the administrator or owner; types of soil, physical constants and profile of soil; types of vegetation and its condition; a description of local topography; a description of surrounding land use; photographs and diagrams of the instrumentation, site and surrounding area; type of AWS, the manufacturer, model and serial number; the observing programme of the station (elements measured, reference time, times at which observations and measurements are made and reported, and the datum level to which atmospheric pressure data of the station refer); and contact information, such as name and mailing address, electronic mail address and telephone numbers.

The documentation should contain a complete history of the station, giving the dates and details of all changes. It should cover the establishment of the station, the commencement of observations, any interruptions to operation, and eventually the station's closure. Comments from inspection visits (see 2.6.6) are also important, especially comments about the site, exposure, quality of observations and station operations.

Instrument metadata should include sensor type, manufacturer, model and serial number; principle of operation; method of measurement and observation; type of detection system; performance characteristics; unit of measurement and measuring range; resolution, accuracy (uncertainty), time constant, time resolution and output averaging time; siting and exposure (location, shielding and height above or below ground); date of installation; data acquisition (sampling interval and averaging interval and type); correction procedures; calibration data and time of calibration; preventive and corrective maintenance (recommended and scheduled maintenance and calibration procedures, including frequency, and a description of procedures); and results of comparison with travelling standards.

For each individual meteorological element, metadata related to procedures for processing observations should include the measuring and observing programme (time of observations, reporting frequency and data output); the data processing method, procedure and algorithm; formulae for calculations; the mode of observation and measurement; the processing interval; the reported resolution; the input source (instrument and element); and constants and parameter values.

Data-handling metadata should include quality control procedures and algorithms, definitions of quality control flags, constants and parameter values, and processing and storage procedures. The transmission-related metadata of interest are method of transmission, data format, transmission time and transmission frequency.

Upper-air stations have metadata requirements that are similar to those of surface stations. In addition, they must maintain metadata on each of the expendable instruments used (such as radiosondes).

Specific guidance is provided in *WIGOS Metadata Standard* (WMO-No. 1192) and reflects the need to specify an observed variable; to answer why, where and how the observation was made, how the raw data were processed, and what the quality of the observation is.

# REFERENCES


World Meteorological Organization, 1983, 1996, 2008, 2014: *Guide to Meteorological Instruments and Methods of Observation* (WMO–No. 8). Fifth, sixth, seventh and eighth editions. Geneva.

———, 1994: *Guide to the Applications of Marine Climatology* (WMO-No. 781). Geneva.

———, 1994: *Observing the World's Environment: Weather, Climate and Water* (WMO-No. 796) (J.P. Bruce). Geneva.

———, 1995: *Plan for the Global Climate Observing System* (GCOS) (WMO/TD-No. 681, GCOS-No. 14), Version 1.0. Geneva.

———, 1995: *GCOS Guide to Satellite Instruments for Climate* (WMO/TD-No. 685, GCOS-No.16). Geneva.

———, 1995: *GCOS Plan for Space-Based Observations* (WMO/TD-No. 684, GCOS-No. 15), Version 1.0. Geneva.

———, 1997: *Report of the GCOS/GOOS/GTOS Joint Data and Information Management Panel*, third session (Tokyo, Japan, 15–18 July 1997) (WMO/TD-No. 847, GCOS-No. 39). Geneva.

———, 1998: *Preliminary Statement of Guidance Regarding How Well Satellite Capabilities Meet WMO User Requirements in Several Application Areas* (WMO/TD-No. 913, SAT-21). Geneva.

———, 1999: *Representativeness, Data Gaps and Uncertainties in Climate Observation*s (lecture given by Chris Folland to the Thirteenth World Meteorological Congress, 21 May 1999) (WMO/TD-No. 977, WCDMP-No. 44). Geneva.

———, 2000: *SPARC Assessment of Upper Tropospheric and Stratospheric Water Vapour* (WMO/TD-No. 1043, WCCRP-No. 113). Geneva.

———, 2000: *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO 2000) – Lightning Detection Systems* (WMO/TD-No. 1028). Geneva.

———, 2000: *Statement of Guidance Regarding How Well Satellite Capabilities Meet WMO User Requirements in Several Application Areas* (SAT-22, WMO/TD-No. 992). Geneva

———, 2002: *Guide to the GCOS Surface and Upper-air Networks*: *GSN and GUAN* (WMO/TD-No. 1106, GCOS-No. 73), Version 1.1. Geneva.

———, 2003: *Report of the Capacity-building Training Workshop on Reducing the Impacts of Climate Extremes on Health* (WCASP-No. 59, WMO/TD-No. 1162). Geneva

———, 2003: *Guidelines on Climate Observation, Network and Systems* (WMO/TD-No. 1185, WCDMP-No. 52). Geneva.

———, 2004: *The Changing Atmosphere: An Integrated Global Atmospheric Chemistry Observation Theme for the IGOS Partnership. GAW Report No. 159* (WMO/TD-No. 1235, ESA SP-1282). Geneva.

———, 2004: *Guidelines on Climate Data Rescue* (WMO/TD-No. 1210, WCDMP-No. 55). Geneva.

———, 2004: *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No. 1188). Geneva.

———, 2004: *Implementation Plan for the Global Climate Observing System for Climate in Support of the UNFCCC* (WMO/TD-No. 1219, GCOS-No. 92). Geneva.

———, 2006: *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites* (WMO/TD-No. 1250). Geneva.

———, 2006: *Systematic Observation Requirements for Satellite-based Products for Climate* (WMO/TD-No. 1338, GCOS-No. 107). Geneva.

———, 2006: *Technical Regulations* (WMO-No. 49), Volume III: Hydrology. Geneva.

———, 2007: *Guidelines for Managing Changes in Climate Observation Programmes* (WMO/TD-No. 1378, WCDMP-No. 62). Geneva.

———, 2008: *Guide to Hydrological Practices* (WMO-No. 168), Volume I: Hydrology – From Measurement to Hydrological Information. Geneva.

———, 2009: *Guide to Hydrological Practices* (WMO-No. 168), Volume II: Management of Water Resources and Application of Hydrological Practices. Geneva.

———, 2009 : *Guideline for the Generation of Satellite-based Datasets and Products Meeting GCOS Requirements* (GCOS-128, WMO/TD-No. 1488). Geneva.

———, 2010: *Guide to the Global Observing System* (WMO-No. 488). Geneva.

———, 2010: *Guide to Agricultural Meteorological Practices*. (WMO-No. 134). Geneva.

———, 2011: *Manual on the Global Observing System* (WMO-No. 544), Volume II: Regional Aspects. Geneva.

———, 2014: *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8), 2017 update. Geneva.

———, 2015: *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160). Geneva.

———, 2015: *Manual on the Global Observing System* (WMO-No. 544), Volume I: Global Aspects. Geneva.

_____, 2016: *Executive Council: Abridged Final Report of the Sixty-Eighth Session* (WMO-No. 1168), Resolution 5 (EC-68) – Competencies for provision of climate services. Geneva.

———, 2016: *Technical Regulations* (WMO-No. 49), *Volume II: M*eteorological Service for International Air Navigation. Geneva.
———, 2017: *Technical Regulations* (WMO-No. 49)*, Volume I: General Meteorological Standards and Recommended Practices.* Geneva.
———, 2017: *WIGOS Metadata Standard* (WMO-No. 1192). Geneva.
———, 2017: *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485). Geneva.

## FURTHER READING

Bénichou, P., 1992: Optimisation de l'implantation d'un réseau complémentaire en zone accidentée: application au réseau automatisé d'Auvergne. *La Météorologie*, 43-44: 3–17.
Christian, H., 2003: Global lightning activity. In: *Proceedings of the 12th International Conference on Atmospheric Electricity*, Versailles, France, 9–11 June 2003 (S. Chauzy and P. Laroche, eds). Paris, ONERA.
Cronin, T.M., 1999: *Principles of Paleoclimatology*. New York, Columbia University Press.
Der Mégréditchian, G., 1985: Méthodes statistiques d'analyse et d'interpolation spatiales des champs météorologiques. *La Météorologie*, 17:51–66.
Dover, J. and L.J. Winans, 2002: Evaluation of windshields for use in the automated surface observing system (ASOS). In: *Proceedings of the Sixth Symposium on Integrated Observing Systems*, Orlando, Florida, 14–17 January 2002. Boston, American Meteorological Society.
Free, M., I. Durre, E. Aguilar, D. Seidel, T.C. Peterson, R.E. Eskridge, J.K. Luers, D. Parker, M. Gordon, J. Lanzante, S. Klein, J. Christy, S. Schroeder, B. Soden and L.M. McMillin, 2002: Creating climate reference datasets: CARDS workshop on adjusting radiosonde temperature data for climate monitoring. *Bulletin of the American Meteorological Society*, 83:891–899.
Free, M. and D.J. Seidel, 2005: Causes of differing temperature trends in radiosonde upper air datasets. *Journal of Geophysical Research*, 110:D07101.
Global Ocean Observing System (GOOS), 2000: *Report of the First GOOS Users' Forum* (GOOS-No. 92). Paris, UNESCO.
———, 2001: *Principles of the Global Ocean Observing System (GOOS) Capacity Building* (GOOS-No. 69). Paris, UNESCO.
———, 2001: *GOOS Capacity Building Implementation Strategy* (GOOS-No. 106). Paris, UNESCO.
Haas, T.C., 1992: Redesigning continental-scale monitoring networks. *Atmospheric Environment*, 26A:3323–3333.
Karl, T.R., C.N. Williams Jr, P.J. Young and W.M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *Journal of Climate and Applied Meteorology*, 25:145–160.
Leroy, M., 1998: Meteorological measurements representativity: nearby obstacles influence. In: *Tenth Symposium on Meteorological Observations and Instrumentation*, Phoenix, Arizona, 11–16 January 1998. Boston, American Meteorological Society.
———, 1998: Climatological site classification. In: T*enth Symposium on Meteorological Observations and Instrumentation*, Phoenix, Arizona, 11–16 January 1998. Boston, American Meteorological Society.
Lieth, H. (ed.), 1974: *Phenology and Seasonality Modeling*. New York, Springer.
National Climatic Data Center (NCDC), 2002: *U.S. Climate Reference Network Site Information Handbook*. Asheville.
Panel on Climate Change Feedbacks, Climate Research Committee, National Research Council, 2003: *Understanding Climate Change Feedbacks*. Washington, DC, National Academy Press.
Panel on Reconciling Temperature Observations, Climate Research Committee, Board on Atmospheric Sciences and Climate, National Research Council, 2000: *Reconciling Observations of Global Temperature Change.* Washington, DC, National Academy Press.
———, 2000: *Improving Atmospheric Temperature Monitoring Capabilities: Letter Report*. Washington, DC, National Academy Press.
Ohring, G., B. Wielicki, R. Spencer, B. Emery and R. Datla, 2005: Satellite instrument calibration for measuring global climate change: Report of a workshop. *Bulletin of the American Meteorological Society*, 86:1303–1313.
Oke, T.R., 1987*: Boundary Layer Climates*. Second edition. Cambridge, Cambridge University Press.
———, 2006: Towards better scientific communication in urban climate. *Theoretical and Applied Climatology*, 84:179-190.

Quadrelli, R. and J.M. Wallace, 2004: A simplified linear framework for interpreting patterns of Northern Hemisphere wintertime climate variability. *Journal of Climate*, 17:3728–3744.

Santer, B.D., M.F. Wehner, T.M.L. Wigley, R. Sausen, G.A. Meehl, K.E. Taylor, C. Ammann, J. Arblaster, W.M. Washington, J.S. Boyle and W. Bruggemann, 2003: Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, 301(5632):479–483.

Schwartz, M.D. (ed.), 2003: *Phenology: An Integrative Environmental Science. Tasks for Vegetation Science*, Volume 39. Dordrecht, Kluwer Academic Publishers.

Seidel, D.J. and M. Free, 2006: Measurement requirements for climate monitoring of upper air temperature derived from reanalysis data. *Journal of Climate*, 19:854–871.

Singleton, F. and E.A. Spackman, 1984: Climatological network design. *The Meteorological Magazine*, 113:77–89.

Sneyers, R., 1973: Sur la densité optimale des réseaux météorologiques. *Archiv für Meteorologie, Geophysik und Bioklimatologie,* Serie B, 21:17–24.

Soden, B.J. and I.M. Held, 2006: An assessment of climate feedbacks in coupled ocean–atmosphere models. *Journal of Climate*, 19:3354–3360.

Thompson, D.W.J. and J.M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, 13:1000–1016.

Thorne, P.W., D.E. Parker, J.R. Christy and C.A. Mears, 2005: Uncertainties in climate trends: Lessons from upper air temperature records. *Bulletin of the American Meteorological Society*, 86:1437–1444.

Thorne, P.W., D.E. Parker, S.F.B. Tett, P.D. Jones, M. McCarthy, H. Coleman and P. Brohan, 2005: Revisiting radiosonde upper air temperatures from 1958 to 2002. *Journal of Geophysical Research*, 110 (D18):D18105.

Unninayar, S. and R.A. Schiffer, 1997: *In-situ Observations for the Global Observing Systems: A Compendium of Requirements and Systems*. Washington, DC, National Aeronautics and Space Administration Office of Mission to Planet Earth, NP-1997(01)-002-GSFC.

Wang, J., D.J. Carlson, D.B. Parsons, T.F.Hock, D. Lauritsen, H.L. Cole, K. Beierle and E. Chamberlain, 2003: Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication. *Geophysical Research Letters*, 30:10.1029/2003GL016985.

World Meteorological Organization, 2003: *Aerosol Measurement Procedures – Guidelines and Recommendations* (WMO/TD-No. 1178, GAW-No. 153). Geneva.

———, 2007: *RA IV Training Seminar on Capacity Building in Climate-related Matters* (WMO/TD-No. 1386, WCDMP-No. 63). Geneva.

———, 2008: *Plant Phenological Guidelines* (WMO/TD-No. 1484, WCDMP-No. 70). Geneva.

———, 2010: *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC* (WMO/TD-No. 1523; GCOS-No. 138), 2010 update. Geneva.

———, 2012: *Manual on Marine Meteorological Services* (WMO-No. 558). Volume I: Global Aspects; Volume II: Regional Aspects. Geneva.

———, 2013: *The GCOS Reference Upper-air Network* (GRUAN) – *Manual* (WIGOS Technical Report No. 2013-02, GCOS-No. 170). Geneva.

———, 2017: *Guide to the WMO Integrated Global Observing System* (WMO-No. 1165). Geneva.

Zhang, H.-M. and L.D. Talley, 1998: Heat and buoyancy budgets and mixing rates in the upper thermocline. *Journal of Physical Oceanography*, 28:1961–1978.

# CHAPTER 3. CLIMATE DATA MANAGEMENT

## 3.1 INTRODUCTION

For thousands of years historians have recorded information about the weather. In the past, however, this information was often based on accounts from other people and was not drawn from the historians' personal observations. Such accounts may have been vague, truncated or affected by memory lapses. This type of weather information was embedded within an immense array of other kinds of information, and much of it is contained in national libraries and archives. Specialized national meteorological archives are a relatively recent phenomenon, with the earliest typically being established during the first half of the twentieth century.

Early records in manuscript form were kept in daily, weekly or monthly journals. Notes were made of extreme or catastrophic events such as high or low temperatures, abnormal wind speeds, excessive rainfall or prolonged drought, dates of frost or freezing, hurricanes and tornadoes. Storms, calms, winds, currents, types of cloud and cloudiness were noted in marine logbooks. Freezing and thawing dates of rivers, lakes and seas, as well as the first and last dates of snowfall, were often recorded as an important part of any journal. Pride of work and accomplishment has always been an important feature in weather observing and recording. The person responsible for an observation who signs or seals the logbook still lends authority to records and serves as the personal source of the recorded history.

Specific journals for the collection and retention of climatological information have been established within the last two or three centuries. Up to the 1940s, the forms developed, printed and used in various countries were often different, and the observations were almost always recorded by hand. Since the 1940s, and especially following the establishment of WMO, standardized forms and procedures have gradually become prevalent, and national meteorological archives have been designated as the storage site for these records.

The quantification of climatological data developed as improved instrumentation facilitated the observation of continuous, as well as discrete, variables and the recording of appropriate values in journals or logbooks. For example, thermometers enabled the systematic recording of quantitative temperature measurements and rain gauges facilitated the measurement of precipitation. The development of clock-driven mechanisms permitted the establishment of intensity and duration values and the recording of these data. Other types of recording instruments provided autographic or analogue records. With each new improvement or addition to the tools of observation, the number of items or variables entered in journals and logbooks increased and specially prepared formats were developed. Even as formats have changed, regularity and consistency, or continuity of the record-keeping, have always been highly desirable. A good chronological record should be kept current and in sequential order. Methodical and careful observation and recording permit easier collection, archiving and subsequent use of the records.

In most countries, manuscript forms were sent periodically to a central location. Until the 1970s, these original forms constituted the bulk of all the holdings of climatological information at most collection centres. These centres may have been a section of the local or national government or the central office of an industry such as mining, agriculture or aviation. Gradually, the climatological data-gathering activities affecting national life were assembled within a concerted programme of observation and collection to serve national and international interests.

Since the late twentieth century, most weather information has been transmitted digitally to centralized national collection centres. As the messages have been intended primarily for operational weather forecasting, it has been common practice to rely on the original observing documents for the creation of the climate record in climate centres around the world. The collection, transmission, processing and storage of operational meteorological data, however, are being dramatically improved by rapid advances in computer technology, and meteorological archives are increasingly being populated with data that have never been recorded on paper. The power and ease of use of computers, the ability to record and transfer information electronically, and the development of international exchange mechanisms such as the Internet have given climatologists new tools to rapidly improve the understanding of climate.

Every effort should be made to obtain, in electronic digital form, a complete collection of all primary observed data. Collection of data electronically at the source allows rapid and automatic control measures to be applied, including error checking, prior to the data's transmission from the observation site. In many cases, the collection of climate data by mail may still be cheaper and more reliable, especially in less technologically advanced regions, but unless the data have been recorded on some form of electronic media prior to postage, they will need to be scanned or digitized centrally. The management of the vast variety of data collected for meteorological or climatological purposes requires a systematic approach that encompasses paper records, microform records and digital records. The framework and specifications for a systematic approach are given in *Climate Data Management System Specifications* (WMO-No. 1131).

This chapter discusses general concepts and considerations for managing climate data. Detailed information and guidance relating to specific topics, such as Climate Data Management System (CDMS) requirements, can be found in the reference material listed in the last section of this chapter.

## 3.2 THE IMPORTANCE AND PURPOSE OF MANAGING DATA

The basic goal of climate data management is to preserve, capture and provide access to climate data and products for use by planners, decision-makers and researchers. Permanent archiving is an important objective. The data management system of a climate archive must provide the information to describe the climate of the domain of interest for which the archive has been established, be it national, regional or global. Data produced from meteorological and climatological networks and various research projects represent a valuable and often unique resource, acquired with substantial expenditure of time, money and effort. Many of the ultimate uses of climate data cannot be foreseen when the data acquisition programmes are being planned, and frequently new applications emerge, long after the information is acquired. The initial utilization of meteorological and related data is often only the first of many applications. Subsequent analysis of the data for many and diverse purposes leads to a significant and ongoing enhancement of the return on the original investment in the data acquisition programmes. The global climate change issue, for example, is stretching the requirements for climate data and data management systems far beyond those originally conceived when the original networks were established. To meet these expanding needs, it is critically important that climate information, both current and historical, be managed in a systematic and comprehensive manner. Conventional meteorological data are now augmented by data from a wide array of new instruments and systems, including satellites, radar systems and other remote-sensing devices, and model data, thus making effective and comprehensive CDMSs essential for modern climate centres.

Climatological data are most useful if they are edited, quality-controlled and stored in a national archive or climate centre and made readily accessible in easy-to-use forms. Although technological innovations are occurring at a rapid pace, many climatological records held by National Meteorological and Hydrological Services (NMHSs) are still in non-digital form. These records must be managed along with the increasing quantity of digital records. A CDMS is an integrated computer-based system that facilitates the effective archiving, management, analysis, delivery and utilization of a wide range of integrated climate data.

The primary goals of database management are to maintain the integrity of the database at all times, and to ensure that the database contains all the data and metadata needed to meet the requirements for which it was established, both now and into the future. Database management systems have revolutionized climate data management by allowing efficient storage, access, conversion and updating of many types of data, and by enhancing data security.

It is essential that both the development of climate databases and the implementation of data management practices take into account the needs and capabilities of existing and future data users. While this requirement may seem intuitive, information that is important for a useful application is sometimes omitted, or data centres commit insufficient resources to checking the quality of data for which users explicitly or implicitly demand high quality. For instance, a database without both current and past weather codes could lead to underestimates of the prevalence of observed phenomena. In all new developments, data managers should attempt to have at least one key data user as part of the project team or to undertake some regular consultative process with user stakeholders to keep abreast of both changes in needs and any issues that user communities may have. Examples of stakeholder communities are those involved in climate prediction, climate change, agriculture, public health, disaster and emergency management, energy, natural resource management, urban planning, finance and insurance.

### 3.2.1 CDMS design

All CDMSs are based on some underlying model of the data. This model design is very important for the quality of the resulting system (see Figure 3.1). An inappropriate model will tend to make the system harder to develop and maintain. In general, a database designed for current meteorological data will allow rapid retrieval of recent data from a large number of stations. By contrast, many climate data applications involve the retrieval of data for one or a few stations over a long period. It is essential to document the overall design and underlying data model of the CDMS to facilitate subsequent extension or modification by computer programmers. Similar considerations apply to a metadata model. Details about data models can be found in the *Guidelines on Climate Data Management* (WMO/TD-No. 1376), Data models in use by CDMSs, pp. 32–34, and in *Climate Data Management System Specifications* (WMO-No. 1131), 4.2.3 and 4.3.

The climate data component represents a wide range of time-series climate data. It extends well beyond what may be thought of as traditional meteorological observations and includes:

(a) Global Climate Observing System (GCOS) Essential Climate Variables (ECVs);

(b) An expanded view of climate metadata that includes metadata on observations and discovery data provenance;

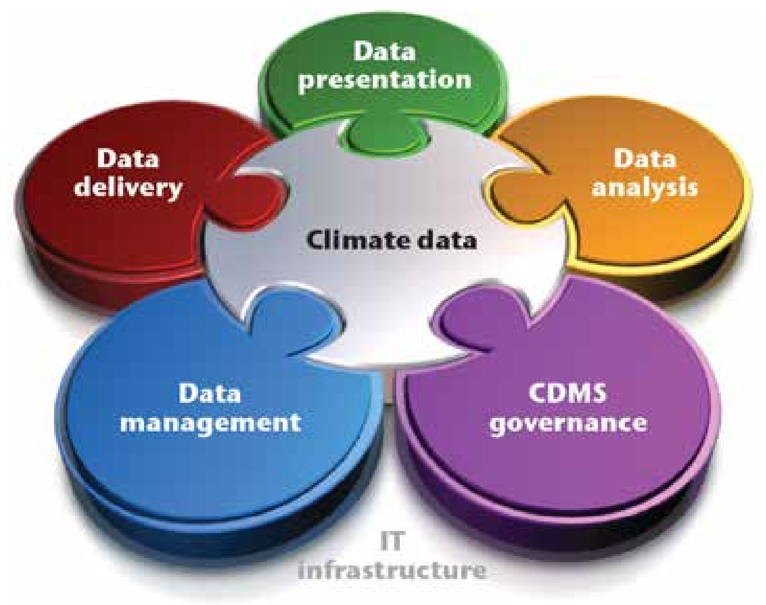



**Figure 3.1. Graphic depiction of the major functional components of a Climate Data Management System**

(c) Standard WMO products;

(d) Derived observations and gridded data;

(e) Outputs from numerical models;

(f) A range of ancillary data used to support CDMSs, including spatial and impact data, documentation and climate software;

(g) Other important data such as logical data models.

The CDMS governance component refers to a consistent set of policies and governance processes needed to build a solid foundation for the establishment and management of authoritative sources of climate data and related services. Although a number of WMO initiatives will establish consistent policies in due course, this component may help to immediately improve many NMHS data management practices. This component contains the following concepts:

(a) Data policy, including organizational commitments, ensuring the sustainability of CDMSs, and intellectual property;

(b) Data delivery;

(c) Third-party data;

(d) Climatology policy;

(e) Governance, including data and IT governance.

The data management component addresses the functionality required to effectively manage climate data and includes the following concepts:

(a) Data ingest and extraction;

(b) Data rescue;

(c) Quality control of observations;

(d) Quality assessment;

(e) Management of climate metadata.

The data delivery component refers to the functionality required to deliver climate data and includes the following concepts:

(a) Data discovery (both climate data and climate metadata);

(b) Data delivery in WMO formats;

(c) Data delivery based on open spatial standards such as the Open Geospatial Consortium (OGC) standards and the International Organization for Standardization (ISO) 19100 series.

The data analysis component involves a wide variety of analytical techniques that are applied to climate data and may result in the generation of a range of derived data products. Some examples are:

(a) A series of techniques, including statistical, spatial and image analysis;

(b) Homogenization;

(c) Numerical modelling processes.

The data presentation component represents a diverse set of techniques used to communicate climate-related information. These include:

(a) Written reports;

(b) Time-series climate data exploration via a graphical user interface with functionalities such as generating a broad variety of business intelligence reports, including tables, graphs, scatter plots, histograms and ensembles; visualizing disparate data using, for example, cartographic techniques, diagrams and 3D; and conducting an integrated search and dynamic exploration of disparate climate data and metadata using functionalities such as spatial intelligence techniques;

(c) Multimedia exploration of data via, for example, podcasts, videos or photographs.

The IT infrastructure components represent the functionalities required to support a CDMS. The components are abstract concepts. A single component may refer to a range of software and processes that provide a functional requirement. Similarly, a single software application may provide the functionalities described in a number of components. The basic CDMS functionality is the same for least developed, developing or developed countries. Not all components are required; *Climate Data Management System Specifications* (WMO-No. 1131) discusses whether a functionality is considered as required, recommended or optional, and provides detailed specifications for all components.

### 3.2.2 **CDMS data acquisition**

Data that are already in digital form can be readily ingested directly by the system. Non-digital records are generally digitized during an entry process. A fundamental goal of a data-entry process is to duplicate, with a minimum of error, the raw data as they were recorded in the capture process. A key-based entry system should be efficient and easy for a data-entry operator to use. The system could also be designed to validate the data as they are entered and to detect likely errors. It is also possible to set default values for some elements, thus saving unnecessary keystrokes.

Where automatic weather stations (AWSs) are in use, climate data, including any error control messages, should be transferred electronically to the CDMS. Manually observed data should be transferred to the CDMS as soon as possible by whatever means are most practical. It is advantageous to collect data at least daily because data quality is likely to be improved, the manual effort for quality control will likely decrease, technical errors will be detected faster, and there will be greater opportunities for improved access to more data. Nevertheless, the submission of data for a month is an acceptable alternative when daily data transmission is not practicable. For example, many of the 6 000 or so voluntary observers in Australia continue to forward monthly rainfall reports that contain the daily observations for the month.

Many weather observations are recorded by institutions or organizations other than NMHSs, and acquisition of the data in their original form may be difficult. In these cases efforts should be made to collect copies of the original report forms. If it is impossible to secure either the original or a copy of the record, a note to this effect should be made in the inventory of the centre’s holdings, stating information pertaining to the existence and location of the data, volume available, period of record covered, stations in the network as applicable and elements observed.

Though not a formal requirement, it is recommended that the CDMS also contain information about media reports, pictures and other similar information beyond the traditional data and metadata. Such information could be captured by imaging the report from the print media with a digital camera or scanner; defining the date, area and type of event (such as flood, drought or heavy precipitation); identifying the media; and writing additional comments about the event.

It is important to retain both the data value that was originally received, as well as the latest quality-controlled value. The original value will likely pass through an initial automated quality control process at ingestion and, as necessary, a more extensive quality control process; even if rejected by either of these processes, it must still be retained. Some CDMSs retain not only the original and latest values, but also all modifications.

Another aspect of data acquisition is the recording of occurrences when data were expected but not received. Loss of data can occur as a result of situations such as inoperable instrumentation, data transmission errors and acquisition processing errors. Lost data can be reconstructed with varying levels of certainty. For example, a missing precipitation measurement can be considered to be zero when it is known from other data that local and synoptic conditions precluded the occurrence of precipitation. In other cases, lost data can be estimated with reasonable certainty using the techniques discussed in 5.9. In all cases, dataset documentation should flag the reconstructed or estimated data appropriately (see 3.4.1).

### 3.2.3 **CDMS data documentation**

An adequate set of metadata must be available to inform future users about the nature of the data in the system, how the various datasets were collected, and any inherent problems. It is recommended that database management include all information that can affect the homogeneity of a dataset or series, including those factors outlined in section 2.6.9.

The structure of metadata in an ideal system is generally more complex than the data themselves. For example, a rainfall observation essentially gives the quantity of precipitation over a certain period at a certain station. The associated metadata that can be applied to this observation, and which might be needed to interpret the data fully, could include information such as the reference date used by the database (for example, time zone); quality indicators or flags that have been ascribed to the observation; history of changes made to the values and any associated flags; instrument used to record the observation, together with details about maintenance programmes, tolerances, internal parameters and similar information; name and contact information of the observer; full details of the location and siting of a station and its history; programme of observations in effect at the time and its history; and topographical and ground-cover details of the site. A detailed treatment of station-specific metadata can be found in the *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186), 2.1–2.5. Similarly, the metadata associated with a gridded dataset of satellite observations of high-resolution solar exposure should include the geographical extent of the observations, the period of record of the dataset, a history of revisions to and maintenance of the dataset, the satellites from which the data were obtained, transfer functions and averaging procedures to obtain grid values, satellite positional accuracy, information about the accuracy of the data, and contact information.

Metadata are also needed for the CDMS itself. Each process within the system (for example, key entry or quality control) should be completely described. A history of any changes made to any part of the system (for example, software, hardware or manual procedures) should be documented and maintained. Since observing practices, quality control techniques and data-handling procedures change over time, these metadata are critical in the climatological analysis of historical data. The analyst uses the metadata to identify and understand how a data value was observed and processed in order to separate meteorological from possible non-meteorological influences in the data record.

Another category of metadata is the record of data holdings in the CDMS. Inventories of the data contained in the CDMS should be prepared routinely. Stratification could be, for example, by data element, station location, or time. Lists of contents should be established and maintained to describe and define the data content of the individual files and to provide information on the codes and observational practices used. Knowing what is contained in the CDMS is important for efficient retrieval of information from the system. The WMO core profile of the ISO 19100 series for data and metadata should be used unless it is superseded by published climate metadata standards.

### 3.2.4 CDMS data storage

An important function of the data manager is to estimate data storage requirements, including the estimation of future growth. Account must be taken of the additional information to be included in data records (for example, data quality flags, original messages, and date and time of record updates), metadata needs and any redundancy necessary to ensure that databases can be restored. Some data types, such as those from remote-sensing, oceanography and AWSs with high temporal resolution, require large amounts of storage.

Unconventional data (such as soil moisture, phenological observations and vegetation indices) may have storage needs that are different from the more traditional observations. Automatic weather stations will often generate data that are relevant to the quality of observations, but are not strictly climate data (for example, information on the battery-level voltage for an AWS). Generally, this information should be utilized prior to data archiving; if it is not included in the CDMS, it should be permanently retained elsewhere and made accessible to data managers. The quality control process often generates values and information that may be different from the original data, so there is a storage requirement for keeping both the original data and any different data generated by quality control processes.

Estimating future growth can be very difficult, as it is hard to determine what data types may become available as time and technology progress. The data manager must consider all of these factors when determining storage requirements.

Non-digital records should be stored in a way that minimizes their deterioration. They should be stored in a controlled environment to avoid temperature and humidity extremes, insects, pests, fire, flood, accidents or deliberate destruction. An ideal example would be storage in acid-free boxes in air-conditioned, secure storerooms. A maintenance programme should be established to rescue deteriorating documents and particularly the data they contain.

With an ever-increasing amount of information being generated and retained, the problem arises as to whether or not to continue to store all the records in their original manuscript form. All too often, climatological records are stored in basements, sheds and other undesirable facilities. They are frequently not catalogued, and they may be inaccessible and subject to deterioration. As a means of reducing paper costs, making better use of space, and providing security for original documents, it is recommended that the manuscript data be scanned into a digital file and carefully preserved. The specifications of computer hardware required for storing and retrieving documents depend on data needs and the limits of financial resources, and also on technological advances, so there are no global standards or a preferred single storage medium. It is important to remember that no storage medium is permanent, and therefore regular review of archival arrangements should be undertaken. Computer-based archives must be securely and regularly backed up with at least one copy stored at a site separate from the main archive.

Microform refers to document images photographically reduced to a very small fraction of their original size. A variety of microform formats exists. Examples are microfilm reels, microfiche sheets, jackets, film folios, aperture cards, cartridges and cassettes. Significant advances in digital storage capabilities, however, now make it highly desirable that paper documents be directly scanned or digitally photographed into a computer system along with superseded microform images. This process facilitates access and ensures preservation for future generations.

### 3.2.5 CDMS data access and retrieval

An important aspect of any CDMS is the power of the facilities to perform data retrieval and analysis. Graphical user interface retrieval facilities should be provided for most users, and command line facilities should be available for the small number of knowledgeable users who have a need for non-standard retrievals. Users should be able to specify their own retrieval criteria, and the system documentation should be clear and provide as much information as necessary to support the users.

Output options should be extensive and include facilities for customizing stations, times and details of output presentations. Users should be given access to listings of data, tabular summaries, statistical analyses and graphical presentation.

### 3.2.6 CDMS archives

An archive is the long-term, and in some cases permanent, means of retention of the data and metadata in the CDMS. The archive structure, whether simple or complex, physical or electronic, should be guided by financial resources, the degree of training of archive personnel, the volume of data to be archived, the media of the data (such as paper documents or digital format), the ease of putting information into and retrieving information from the archive, user-friendliness in accessing information, the ease of maintenance of the archive, and the ease of expansion as data holdings increase. All aspects of the CDMS should be archived, including not just the data values but also catalogues, inventories, histories, dictionaries, and similar information.

### 3.2.7 CDMS security

The main goal of a security policy and associated activities is to prevent loss of or damage to the CDMS. To achieve this goal, the requisites are:

(a) All personnel must be aware of their professional responsibilities;

(b) The archives and database environment must be secured and protected against physical hazards to the records, such as fire and excess humidity;

(c) For digital data, user-level security should be enforced with respect to the database and its components. Only a small and registered group of people should have the right to perform data manipulations such as insertions, updates or deletions;

(d) Personnel with write access to a database must agree not to perform any transactions besides the operations and practices approved by the data manager;

(e) All changes to data tables should have an audit trail, and controls on access to this trail should be in place;

(f) Password security principles should be applied, including not sharing passwords, not writing passwords on paper, changing passwords regularly, and using “strong” passwords consisting of seemingly unrelated letters, numbers and characters;

(g) All unnecessary services should be disabled on the database computer;

(h) The database must be protected against attacks from viruses and hackers;

(i) Regular backups must be made, noting that work done after the most recent backup will likely be lost and need to be repeated should a computer failure occur. Typically, an incremental backup should be made daily and a full backup weekly;

(j) Every so often, typically monthly, a complete backup of the data tables should be put in a safe, secure, fireproof location, remote from the physical location of the climate database. It is common to have three copies of the same archive in different secure places and if possible in different towns or cities;

(k) Backups of the CDMS must be performed prior to any changes to the system software, to the system design or to the applications contained in the CDMS.

### 3.2.8 CDMS management

A CDMS should be monitored routinely to determine how well the processes that use and support the database are performing. Examples of the processes that support the data are metadata maintenance, database ingestion, quality control actions that modify the database, and information retrieval. Each process should be monitored, evaluated and, if necessary, improved. It is strongly recommended that data managers think in terms of end-to-end data management, with information on systemic data quality issues, loss of data, or other practices that harm the climate record being referred back to observation managers for rectification.

Typical monitoring reports would include the number and type of stations in the database, the quantity of data in the database grouped by stations and by observation element types, and information about missing data. This information can be compared to observation schedules to identify when and where data are being lost. Other reports could include quality control actions to ensure that the quality control process is performing properly with new data or to identify any groupings of data with excessive quality problems. It is useful to track the quantity and range of data being retrieved for user inquiries, since this information is helpful in identifying both the most important datasets and the areas to be developed in the future.

The frequency and reporting period of monitoring reports depend on the needs of the NMHS. Reports on data ingestion may be made automatically, perhaps every day. Monthly reports on the quantity and quality of data usually match the monthly cycle of many climate products.

### 3.2.9 International CDMS standards and guidelines

There is no agreement on the optimal structure of a climatological database, as the design depends on the specific needs of the NMHSs and stakeholders. One need may be to access all data for a specified element over a region for a given time, but another need may be to access a data time series for the same element for a single location. The particular needs will have a strong impact on the required storage space or the response time to load or to access data. General principles that should be followed in any design, however, include:

(a) User documentation: Manuals should include an overview of the database and instructions for installation, with detailed information for users, system administrators and programmers;

(b) Key entry: On-screen layouts of data input forms should be similar to the layout of the paper forms from which the data are being copied; customization of layouts should be possible; procedures for entering information into the database should meet the needs of the NMHS; validation of entries (for example, permissible values or station identifiers) should be automatically performed in the key entry process, and default values should be entered automatically;

(c) Ingestion of digital data: The system should be able to automatically ingest standard formatted data that are transmitted via the Global Telecommunication System (GTS), files containing data for multiple stations, multiple files containing data for a single station, data from AWSs, data with user-defined formats, and metadata;

(d) Validation and quality control: The system should provide flags indicating the data source, level of quality assurance performed (such as key entry process or end-of-month processing), quality assurance results, and the reason for the decision to accept, reject or estimate a value. It should retain original, estimated and changed data values; it should also evaluate data temporally and spatially for permissible values, meteorological consistency and physical reasonableness;

(e) Technical documentation: There must be listings defining each table in the database and the relationships among tables; naming conventions should be consistent across all tables, indexes, entities and views;

(f) Data access: The interface between the user and the database should be easy to use, and procedures for extracting information should be documented with clear instructions and examples;

(g) Metadata: The system must be able to manage the full range of metadata (as described in 2.6.9);

(h) Outputs: CDMSs should be able to produce standard outputs that meet the needs of the NMHS, such as data listings and tabulations of hourly, daily, monthly and longer-period data; statistical summaries; and graphical depictions such as contour analyses, wind roses, time series, upper-air soundings and station model plots;

(i) Data and system administration: The system should allow for routine and hoc back up without being shut down; restoration as needed in a timely manner; logging of individual transactions; maintenance of security; monitoring for system performance (for example, memory usage, available storage space, number of transactions and status of system logs); and copying at regular intervals to a separate physical location;

(j) Level of assistance: Users should be able to solve problems using available documentation, interact with other users to exchange questions and comments, and obtain advice from developers of the system as needed and in a timely manner;

(k) Flexibility: The system should be capable of expansion and modification as hardware and software technologies evolve and data sources change, and as the need for output products increases.

## 3.3 QUALITY CONTROL

The objective of quality control is to verify whether a reported data value is representative of what was intended to be measured and has not been contaminated by unrelated factors. It is important, therefore, to be clear from the outset what the readings of a particular data series are meant to represent. Data should be considered as satisfactory for permanent archiving only after they have been subjected to adequate quality control.

The observer or automated observing system should apply quality control to ensure that the time and station identification are correct, that the recorded values reliably reflect current conditions, and that there is consistency among the observed elements. These steps must all be taken prior to the recording or transmission of an observation.

The archiving centre should also apply quality control to the observations received. If manuscript records constitute the source document, trained personnel should, upon receipt at the archiving centre, scrutinize them before any digitization takes place. The forms should be reviewed to ensure proper identification (for example, station name, identifying number and location), legibility and proper recording of data (for example, to the correct precision and in the proper columns). If any problems are discovered, the observation sites should be contacted to clarify the issues or correct the problems. If resources do not permit the quality control of all data, priority should be given to the most important climate elements.

Procedures for quality control of meteorological data help assure defined data quality levels throughout the life cycle of the meteorological data and should be an integral part of an entity’s quality management system. It should be noted that the procedures serve in particular to assure data quality levels that support climate applications and services. It should also be noted that quality control procedures need to be adapted to the specific climate conditions of a country and fine-tuned to fit existing and planned observational and information technology infrastructure as well as available human resources. Applied quality control procedures should be well documented and made available to data users. Detailed guidelines for quality assurance of surface data for climate applications will be given in a forthcoming replacement of *Guidelines on the Quality Control of Surface Climatological Data* (WMO/TD-No. 111). The new publication will include a number of tests that will augment the information contained in 3.4.3–3.4.7 below.

### 3.3.1 Quality control procedures

When observed data are available in digital form, the archiving centre should subject them to full, elaborate quality control procedures on a regular, systematic basis. Computer programs can examine all the available data and list those that fail pre-defined tests, but are not so adept at identifying the underlying problem. A skilled human analyst can often make judgments about the cause of errors and determine any corrections that should be applied, but is generally overwhelmed by the vast quantity of observations. The best technique is a combination of the two, with computer-generated lists of potential errors presented to the human analyst for further action.

Statistical techniques (described in Chapters 4 and 5) are invaluable for detecting errors, and in some cases for suggesting what the "correct" value should be. Objective, automated screening of data is essential when validating large quantities of data. A manual review of the automated output is needed, however, to ensure that the automated procedures are indeed performing as expected. Graphical and map displays of data and data summaries are excellent tools for visual examinations. These techniques integrate and assimilate large quantities of data and enable a trained analyst to recognize patterns for assessing physical reasonableness, identifying outliers, noticing suspect data and evaluating the performance of automated procedures.

All observations should be appropriately flagged. Corrections or estimated correct data should be entered into the database. The original data, however, must also be retained in the database. After the data have been quality-controlled, corrected and edited, the final dataset once again should be put through the quality control checks. This last step will help ensure that errors have not been introduced during the quality control procedures. Further manual review should help identify patterns of errors that may have resulted from, for example, software errors, and inadequate or improper adherence to instructions or procedures. The patterns should be relayed to managers of the NMHS observation program.

In a database a given value is generally available at different stages of quality control. The original data as received in the database must be kept, but validation processes often lead to modifications of the data. These different stages of the value are reflected in quality flags. A multitude of flags could be constructed, but the number of flags should be kept to the minimum needed to describe the quality assessment and reliability of the raw data or estimated values. A quality flag code using two digits, one for the type of data and one for the stage of validation, meets most requirements. When data are acquired from multiple sources, a third flag for the source of the data is often useful. Examples of "type of data", "validation stage" and "acquisition method" codes are given in Tables 3.1, 3.2 and 3.3.

**Table 3.1. Example of type of data codes**

| *Type of data code* | *Meaning* |
|---|---|
| 0 | Original data |
| 1 | Corrected data |
| 2 | Reconstructed (such as by interpolation, estimation or disaggregation) |
| 3 | Calculated value |

**Table 3.2. Example of a stage of validation flag code**

| *Validation stage code* | *Meaning* |
|---|---|
| 1 | Missing data (data not received or observation not made) |
| 2 | Data eliminated once controls completed |
| 3 | Not controlled (newly inserted data or historical data not subject to any control) |
| 4 | Declared doubtful as identified as an outlier by preliminary checks, awaiting controls (data possibly wrong) |
| 5 | Declared doubtful after automatic controls or human supervision (data probably wrong) |
| 6 | Declared validated after automatic controls or human supervision (but further modification allowed, for example, if a subsequent study reveals that the data can still be improved) |
| 7 | Declared validated after automatic controls and human supervision and no further modification allowed |

**Table 3.3. Example of a data acquisition method flag code**

| *Acquisition method code* | *Meaning* |
|---|---|
| 1 | Global Telecommunication System |
| 2 | Key entry |
| 3 | Automated weather station telecommunication network |
| 4 | Automated weather station digital file |
| 5 | Manuscript record |

### 3.3.2 Quality control documentation

Quality control procedures and algorithms should be documented in detail for each stage of data processing from observation to archiving. The checks that are performed by the observer, initial validation by the collection centre, final validation, quality control of changed formats for archiving or publication, and checks on summarized data all require documentation.

Detailed records and documentation should be accessible to users of the data. Knowledge of the data-processing and quality control procedures allows a user of the data to assess the validity of the observation. With proper documentation and retention of the original data, future users are able to assess the impact of changes in procedures on the validity, continuity or homogeneity of the data record; apply new knowledge in atmospheric science to the older data; and perhaps revalidate the data based on new techniques and discoveries.

### 3.3.3 Types of error

Metadata errors often manifest themselves as data errors. For example, an incorrect station identifier may mean that data from one location apparently came from another; an incorrect date stamp may mean the data appear to have been observed at a different time. Data that are missing for the correct place and time should be detected by completeness tests; data that have been ascribed to an incorrect place or time should be detected by consistency and tolerance tests.

Data errors arise primarily as a result of instrumental, observer, data transmission, key entry and data validation process errors, as well as changing data formats and data summarization problems. When establishing a set of quality control procedures, all potential types, sources and causes of error should be considered and efforts should be made to reduce them. It is recommended that, in developing automated and semi-automated error-flagging procedures, system designers work closely with operational quality control personnel.

### 3.3.4 **Format tests**

Checks should be made for repeated observations or impossible format codes such as alpha characters in a numeric field, embedded or blank fields within an observation, impossible identification codes, and impossible dates. The actual causes of a format error could include miskeying, the garbling of a message in transmission, or a mistake by an operator.

Procedures should be introduced to eliminate, or at least reduce, format errors. Two methods commonly employed that reduce key entry errors are double entry (where the same data are entered independently by two operators) and error detection algorithms. Which method is better depends on the skills of the data-entry personnel, the complexity of the observation and the resources available. Digital error detection and correction techniques should be used to eliminate, or at least detect, transmission errors. Careful design of data-entry systems can minimize operator errors, but proper training and performance checks are needed even with the most user-friendly system.

### 3.3.5 **Completeness tests**

For some elements, missing data are much more critical than for others. For monthly extremes or event data such as the number of days with precipitation greater than a certain threshold, missing daily data may render the recorded value highly questionable. Total monthly rainfall amounts may also be strongly compromised by a few days of missing data, particularly when a rain event occurred during the missing period. On the other hand, monthly averaged temperature may be less susceptible to missing data than the two previous examples. For some applications data completeness is a necessity.

Data should be sorted by type of observation into a prescribed chronological order by station. An inventory should be compared to a master station identifier file. Comparison should be made between the observations actually received and the observations that are expected to be received. The absence of any expected observation should be flagged for future review.

### 3.3.6 **Consistency tests**

The four primary types of consistency checks are internal, temporal, spatial and summarization. Since data values are interrelated in time and space, an integrated procedure should be developed to examine consistency. All consistency tests should be completely documented with procedures, formulae and decision criteria.

Internal consistency relies on the physical relationships among climatological elements. All elements should be thoroughly verified against any associated elements within each observation. For example, psychrometric data should be checked to ensure that the reported dry bulb temperature equals or exceeds the reported wet bulb temperature. Similarly, the relationship between visibility and present weather should be checked for adherence to standard observation practices.

Data should be checked for consistency with definitions. For example, a maximum value must be equal to or higher than a minimum value. Physical bounds provide rules for further internal consistency checks. For example, sunshine duration is limited by the duration of the day, global radiation cannot be greater than the irradiance at the top of the atmosphere, wind direction must be between 0° and 360°, and precipitation cannot be negative.

Temporal consistency tests the variation of an element in time. Many climatological datasets show significant serial correlation. A check should be made by comparing the prior and subsequent observations with the one in question. Using experience or analytical or statistical methodologies, data reviewers can establish the amount of change that might be expected in a particular element in any time interval. This change usually depends on the element, season, location and time lag between two successive observations. For example, a temperature drop of 10°C within one hour may be suspect, but could be quite realistic if associated with the passage of a cold front or onset of a sea breeze. The suspicious value will have to be compared to present weather at that time, and perhaps to other types of observations (such as wind direction or satellite, radar or lightning detection) before a decision is made to validate or modify it. For some elements, a lack of change could indicate an error. For example, a series of identical wind speeds may indicate a problem with the anemometer.

Temporal consistency checks can be automated easily. Section 5.5 describes some of the techniques of time series analysis that can be adapted for quality control purposes. Graphical displays of data are also excellent tools for verification. Several elements should be visualized at the same time in order to facilitate diagnostics. For example, it will be easier to validate a temperature drop if the information showing the veering of winds associated with the passage of a cold front, or heavy rain from a thunderstorm, is also available.

Spatial consistency compares each observation with observations taken at the same time at other stations in the area. Each observation can be compared to what would be expected at that site based on the observations from neighbouring stations. Those data for which there is a significant difference between the expected and actual observations should be flagged for review, correction or deletion as necessary. It is important to recognize that only like quantities should be directly compared, such as wind speeds measured at the same height; values measured at similar elevations, such as flat, open topography; or values measured within a climatologically similar area. Section 5.9 details the data estimation techniques that are required for this type of quality control process.

Summarization tests are among the easiest to perform. By comparing different summaries of data, errors in individual values or in each summary can be detected. For example, the sums and means of daily values can be calculated for various periods such as weeks, months or years. Checking that the total of the twelve monthly reported sums equals the sum of the individual daily values for a year provides a quick and simple cross-check for an accumulation element like rainfall. Systematic errors in upper-air station data can sometimes be revealed by comparing monthly averages with the averages derived for the same location and height from a numerical analysis system. The cause of any inconsistencies should be reviewed and corrected.

Marine observations, in general, can be subjected to procedures similar to those used for surface land stations, with slight modification for additional elements, assuming that there is a ship identifier present in each observation to allow data to be sorted chronologically and by ship order. Upper-air observations must be verified in a somewhat different manner. Some cross-checks should be made of surface-level conditions with those at a nearby or co-located surface station. A quality control programme to check upper-air data should compute successive level data from the preceding level starting with the surface data. Limits on the difference allowed between the computed and reported values should be established. Any level with reported elements that fail the test should be flagged as suspect, reviewed or corrected.

### 3.3.7 **Tolerance tests**

Tolerance tests set upper or lower limits to the possible values of a climatological element (such as wind direction, cloud cover, and past and present weather) or, in other cases, where the theoretical range of values is infinite, the limits outside of which it is unlikely for a measurement to lie. In the latter case, the limits are usually time- and location-dependent and should be established by recourse to the historical values or by spatial interpolation methods. It is also important to identify and then quickly address systematic biases in the outputs from instrumentation. Documentation must be maintained regarding which tolerance tests have been applied, the climate limits established for each inspected element, and the rationale for determining these limits.

In general, tolerance tests compare a value in question against some standard using a statistical threshold. Some simple tolerance tests include comparing an observed value to the extreme or record value or to some multiple of standard deviations around the average value for that date. In the latter case, one must take into consideration the possibility that the element may not necessarily have a symmetrical or Gaussian distribution, and that some extreme values identified from the standard deviation multiplier may be incorrect.

When using long-term historical data for quality control, it is preferable to use a standardized reference (for example, standard deviations or a non-parametric rank order statistic) rather than an absolute reference. Section 4.4 discusses the various summary descriptors of data, including the restrictions on their appropriateness.

It may be possible to perform some tolerance tests using completely different data streams, such as satellite or radar data. For example, a very simple test for the occurrence or non-occurrence of precipitation using satellite data would be to check for the presence of clouds in a satellite image.

## 3.4 EXCHANGE OF CLIMATIC DATA

Exchange of data is essential for climatology. For WMO Members, the obligation to share data and metadata with other Members, and the conditions under which these may be passed to third parties, are covered under Resolution 60 of the Seventeenth World Meteorological Congress (with regard to the Global Framework for Climate Services), Resolution 25 of the Thirteenth World Meteorological Congress (for hydrological data), Resolution 40 of the Twelfth World Meteorological Congress (with regard to meteorological data), and Intergovernmental Oceanographic Commission Resolution XXII-6 (for oceanographic data). These resolutions address the concepts of “essential” and “additional” data, with a specification of a minimum set of data that should be made available in a non-discriminatory manner and at a charge of no more than the cost of reproduction and delivery, without requiring payment for the data and products themselves. Members may decide to declare as “essential” more than the minimum set. The use of agreed-upon international standard formats for data exchange is critical.

Beyond CLIMAT and related messages (see 4.8), Members are also asked to provide additional data and products that are needed to sustain WMO programmes at the global, regional and national levels and to assist other Members in providing meteorological and climatological services in their countries. Members supplying such additional data and products may place conditions on their re-export. Research and educational communities should be provided with free and unrestricted access to all data and products exchanged under the auspices of WMO for their non-commercial activities.

Members of WMO voluntarily nominate subsets of their stations to be parts of various networks, including the Global Climate Observing System (GCOS) Upper-air Network (GUAN), the GCOS Surface Network (GSN), the Regional Basic Synoptic Network and the Regional Basic Climatological Network. Nomination of stations to participate in these networks implies an obligation to share the data internationally.

Data are also shared through the International Science Council (ISC) World Data System (WDS). The World Data System works to guarantee access to solar, geophysical and related environmental data. It serves the whole scientific community by assembling, scrutinizing, organizing and disseminating data and information. The system collects, documents and archives measurements and the associated metadata from stations worldwide, and makes these data freely available to the scientific community. In some cases, WDS also provides additional products, including data analyses, maps of data distributions, and data summaries. The World Data System covers meteorology, paleoclimatology, oceanography, atmospheric trace gases, glaciology, soils, marine geology and geophysics, sunspots, solar activity, solar–terrestrial physics, airglow, aurora, and cosmic rays, as well as other disciplines.

The World Meteorological Organization is actively involved in the provision of data to WDS, and there are a number of associated centres operated directly through WMO. The WMO centres deal with ozone and ultraviolet radiation, greenhouse gases, aerosols, aerosol optical depth, radiation and precipitation chemistry. There are differences in data access policy for ISC and WMO centres. International Science Council data centres exchange data among themselves without charge and provide data to scientists in any country free of charge. Data centres operated through WMO must abide by Resolutions 40 and 25 referred to above, which allow for some data or products to be placed in the WDS with conditions attached to their use.

In addition to the International Science Council WDS, there are many centres that operate under cooperative agreements with WMO or with individual NMHSs. These centres include the Global Precipitation Climatology Centre and the Global Runoff Data Centre (Germany); Australia's National Climate Centre; the World Ozone and Ultraviolet Radiation Data Centre (Canada); the Met Office Hadley Centre (United Kingdom); and in the United States, the Lamont-Doherty Earth Observatory of Columbia University, the National Centers for Environmental Information, the National Aeronautics and Space Administration (NASA) Goddard Distributed Active Archive Center, the Tropical Atmosphere Ocean (TAO)/Triangle Trans-Ocean Buoy Network (TRITON) Array, and the University Corporation for Atmospheric Research.

Exchange of digital data is simple for many Members because of the range of computer communication systems available. The Global Telecommunication System is a meteorological communication system with connections to virtually all countries of the world. As an operational system with a critical role in global weather forecasting, it provides reliable communication services, albeit sometimes with low bandwidth. Like the Internet, the Global Telecommunication System is based on a confederation of interconnected networks. As a closed system, however, it is free from the security breaches that often plague the Internet. Open communication linkages such as the Internet should be protected by the best available security software systems to minimize the danger of unwanted access and file manipulation or corruption.

It is highly unlikely that archived formats used for climatological data by one country would be the same as those used by another. The format documentation describing the data organization, element types, units and any other pertinent information should accompany the data. In addition, if the digital data are compacted or in a special non-text format, it is extremely useful for the contributing archive centre to provide "read" routines to accompany digital data requested from an archive.

International data exchange agreements allow for the global compilation of publications on climate normals such as *World Weather Records* and *Monthly Climatic Data for the World*. Bilateral or multilateral agreements are also important in creating and exchanging long-term datasets, such as the Global Historical Climate Network, the Comprehensive Aerological Reference Data Set and the Comprehensive Ocean–Atmosphere Data Sets, compiled by the United States, and the Hadley Centre global observation datasets compiled by the United Kingdom. These datasets are generally provided to research centres.

The current WMO information systems have been developed to meet a diverse set of needs for many different programmes and commissions. The multiplicity of systems has resulted in incompatibilities, inefficiencies, duplication of effort and higher overall costs for Members. An alternative approach designed to improve efficiency of the transfer of data and information among countries is the WMO Information System (WIS). It is envisioned that this system will be used for the collection and sharing of information for all WMO and related international programmes. Non-meteorological and non-climatic environmental and geophysical data such as ecological, earthquake and tsunami data could be included. The WIS vision provides guidance for the orderly evolution of existing systems into an integrated system that efficiently meets the international environmental information requirements of Members.

The WMO Information System will provide an integrated approach to routine collection and automated dissemination of observed data and products, timely delivery of data and products, and requests for data and products. It should be reliable, cost-effective and affordable for developing as well as developed countries. It should also be technologically sustainable and appropriate to local expertise, modular, scalable, flexible and extensible. It should be able to adjust to changing requirements, allow dissemination of products from diverse data sources, and allow participants to collaborate at levels appropriate to their responsibilities and budgetary resources. The WMO Information System should also support different user groups and access policies such as those outlined in Resolutions 40 and 25 referred to above, data as well as network security, and integration of diverse datasets.

## 3.5 DATA RESCUE

Data rescue involves organizing and preserving climate data at risk of being lost due to deterioration, destruction, neglect, technical obsolescence or simple dispersion of climate data assets over time. Non-digitized data are at risk, owing to the vulnerability of the original paper record. Data rescue includes: organizing and imaging paper, microfilm and microfiche records; keying numerical and textual data and digitizing strip-chart data into a usable format; and archiving data, metadata and quality-control outcomes and procedures. An overview of components of climate data rescue activities is provided in Table 3.4.

**Table 3.4. Components of climate data rescue activities**

| *Components of data rescue* | *Activities* | *Keywords* |
|---|---|---|
| Archiving paper and microfilm/microfiche media | Search and locate | NMHSs, observation sites, universities, aviation and maritime agencies, agricultural organizations, international libraries and databases, national archives |
| | Preserve and store | Clean media; place in acid-free, labelled archive boxes, safe from dust, moisture and pests |
| | Create an electronic inventory of paper/microfilm holdings | Catalogue all paper media; estimate scope of imaging and digitizing effort |
| Imaging media | Create a master image inventory | |
| | Image and validate<br>Update the master image inventory<br>Create image file inventories on each CD/DVD or in each computer directory | Update the master image inventory after images have been validated as readable, including metadata<br>Cross-check CD/DVD file inventories with master image inventory |
| Digitizing data values | Create a digital data inventory | |
| | Key entry, chart trace | Data input into the Climate Data Management System |
| | Quality-control data | Update digital data inventory as data are digitized and pass through various quality control tests |
| Archiving digital media | Cross-check printed media, images and digital data | Compare image and digital data inventories with original electronic inventory of paper/microfilm holdings |
| | Back up electronic media | Daily |
| | Disperse multiple copies of images and digital data archives | To various locations |
| | Refresh media and migrate technologies | Every 5 to 10 years |

Any group or individual who has data (paper, microfilm or digital) should attempt to facilitate climate data rescue. Those who are responsible for the management of a country’s climate record should have a special role in data rescue, since they are in a better position to appreciate and value the data being rescued and to know which are most important. These staff are often located in the climate sections of NMHSs, and data rescue is a component of their custodian responsibilities as climate data managers. However, data rescue proponents can be found in many institutions, both public (agricultural departments) and private (universities, plantations, agribusiness). In addition, volunteer organizations, such as the International Environmental Data Rescue Organization (IEDRO) and the International Atmospheric Circulation Reconstructions over the Earth (ACRE) have been created to facilitate data rescue. Other parties interested in assisting with data rescue activities include working and retired climatologists, librarians, historians and students. Such people would also be valuable in implementing crowdsourcing operations.

*Guidelines on Best Practices for Climate Data Rescue* (WMO-No. 1182) provides detailed guidance for all facets of climate data rescue. The International Data Rescue (I-DARE) portal is a web-based resource for people interested in data preservation, rescue and digitization. It provides a single point of entry for information on the status of past and present data rescue projects worldwide, on data that needs to be rescued and on the methods and technologies involved. It is a gateway for the exchange of information on all aspects of data rescue, including established and emerging rescue technologies. Because its goals are to enhance the visibility of existing data rescue activities, to stimulate new ones and to better coordinate international data rescue efforts, the I-DARE portal is a useful communication tool. It will also assist in identifying gaps and opportunities, help prioritize data rescue in regions where it is most needed and aid in attracting funding for projects.

The reorganizing, imaging, digitizing and quality-control steps of rescuing data can be expensive and time-consuming. Beginning and sustaining a data rescue programme begins with the NHMS, which needs to make a clear case for the need for data rescue. This could include the identification of practical questions that can be answered by improved climate data. For instance, historical climate data would allow the NHMS to answer questions regarding the frequency of droughts or heavy rainfall events, or the direction of the strongest wind over the past 100 years. Such information provides a "value-added" product that NMHS managers could charge for, bringing in additional revenue for the NMHS. Moreover, organizing climate data could result in reclaiming hundreds of square metres of NMHS floor space without throwing out valuable historical records. Outside support over a number of years is often required for data rescue activities. Individuals and groups who have already undertaken data rescue may be able to provide advice and argue for funding data rescue efforts.

## REFERENCES


World Meteorological Organization, 1986: *Guidelines on the Quality Control of Surface Climatological Data* (WMO/TD-No. 111). Geneva.

———, 2003: *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186, WCDMP-No. 53). Geneva.

———, 2007: *Guidelines on Climate Data Management* (WMO/TD-No. 1376, WCDMP-No. 56). Geneva.

———, 2014: *Climate Data Management System Specifications* (WMO-No. 1131). Geneva.

———, 2015: *Seventeenth World Meteorological Congress: Abridged Final Report with Resolutions* (WMO-No. 1157). Geneva.

———, 2016: *Guidelines for Best Practices for Climate Data Rescue* (WMO-No. 1182). Geneva.


## FURTHER READING


Cornford, D., 1998: An overview of interpolation. In: *Seminar on Data Spatial Distribution in Meteorology and Climatology* (M. Bindi and B. Gozzini, eds). Volterra, European Union COST Action 79.

Cravatte, S., W. S. Kessler, N. Smith, S. E. Wijffels et al., 2016: *First Report of TPOS 2020*. GOOS-215, available at http://tpos2020.org/first-report/

De Gaetano, A.T., 1997: A quality control procedure for hourly wind data. *Journal of Atmospheric and Oceanic Technology*, 14:137–151.

Graybeal, D.Y., A.T. De Gaetano and K.L. Eggleston, 2004: Complex quality assurance of historical hourly surface airways meteorological data. *Journal of Atmospheric and Oceanic Technology*, 21:1156–1169.

Hungarian Meteorological Service, 1997: *Proceedings of the First Seminar for Homogenization of Surface Climatological Data* (Budapest, 6–12 October 1996). Budapest.

International Organization for Standardization, 2015: *ISO 9000:2015 Quality management systems – Fundamentals and vocabulary.*

———, 2015: *ISO/IEC 17021-1:2015 Conformity Assessment – Requirements for Bodies Providing Audit and Certification of Management Systems*.

———, 2018: *ISO 9004:2018 Quality Management – Quality of an Organization – Guidance to Achieve Sustained Success.*

Kresse, W. and K. Fadaie, 2004: *ISO Standards for Geographic Information*. Berlin, Springer.

Merlier, C., 2001: *Interpolation des données spatiales en climatologie et conception optimale des réseaux climatologiques.* Annex to the report by Météo-France on its activities related to the WMO Commission for Climatology (WMO report – Commission for Climatology Lead Rapporteur on Statistical Methods with Emphasis on Spatial Interpolation).

# CHAPTER 4. CHARACTERIZING CLIMATE FROM DATASETS

## 4.1 INTRODUCTION

Each year more and more data are added to the climatological archives of National Meteorological and Hydrological Services (NMHSs). Climatologists must be able to express all the relevant information contained in the data by means of comparatively few values derived through the application of a range of statistical methods. When chosen and applied with care, statistical processes can isolate and bring to the fore relevant information contained in the data.

This chapter concentrates on descriptive statistics, the tool used to reduce the properties of an otherwise large amount of data to a comprehensible form. Many of the methods described in this chapter are best used with computers to process and display the data. It is necessary, though, to draw attention to the dangers of an overly mechanical application of automated methods of analysis, because of the ease with which automated procedures can be misused and the results misinterpreted. While there are unquestionable advantages to using a computer, assumptions implicit in most analysis software run the risk of being ignored or not clearly articulated, hence leading potentially to erroneous results.

Chapter 5 concentrates on statistical methods and should be used in conjunction with this chapter. Both chapters are intended to describe basic concepts rather than to provide detailed specifics of complex subjects. The references at the end of the chapter and textbooks on statistical theory and methods provide more detailed information.

## 4.2 DATASET EVALUATION

A dataset consists of a collection of observations of elements. An observation is a single estimate of some quantity. Simple observations include reading a thermometer or the level of water in a rain gauge. Other observations are more complex. For example, obtaining barometric pressure from a mercury barometer involves taking observations of both the length of the column of mercury and the temperature of the barometer. The pressure is regarded as a single observation.

Some elements are continuous, that is, there are no discontinuities in the state of the phenomenon observed, such as air temperature. Some elements such as precipitation are not continuous while others do not have quantitative values but only a descriptive category, such as a cloud type or present weather description. The population is all possible values for an element. If an element is continuous, then the population is theoretically infinite. If the element is not continuous, then the population is all the specific values that the element can have within boundaries defined by the analyst.

A sample is a set of observations from the population, which is taken to represent the entire population. Datasets are samples. The larger the sample size, the more accurate the estimate of the descriptive features of the population will be. Much of climatology is concerned with the study of samples, but the analyst must recognize that a dataset may be representative of only a part of the population. The influence of, for example, inhomogeneities, dependence on time, and variations in space complicate the interpretation of what the dataset represents.

Prior to the description or use of a dataset, the data should be checked for accuracy and validity. Accuracy refers to the correctness of the data, while validity refers to the applicability of the data to the purpose for which the values will be used. The user of a dataset should never assume without confirmation that a dataset is accurate and valid, especially without relevant information from the quality control processes applied during the assembling of the dataset. It is also important to know how the data have been collected, processed and compiled, and sometimes even to know why the data were initially collected. Chapter 3 covers climate data management, with a discussion of the importance of metadata and quality control issues in sections 3.2 and 3.3, respectively.

## 4.3 QUALITATIVE VISUAL DISPLAYS OF DATA

Some of the basic features of a dataset that are often sought are the middle or typical value, the spread or range of the observations, the existence of unexpected observations, how the observations trail off from either side of the middle value, and the clustering of observations. Without systematic organization, large quantities of data cannot be easily interpreted to find these and similar features. The first step is to gain a general understanding of the data through visual displays of the distribution of the observed values.

There are many ways of portraying data to obtain a qualitative appreciation of what the data are telling the climatologist. One way to organize a dataset is to sort the observations by increasing or decreasing magnitude. The ordered observations can then be displayed graphically or as a table, from which some characteristics, such as extreme values and the range, become apparent.

A second way to organize a dataset is to group the data into intervals. Counts are made of the number of observations in each interval. A graphical display of the number of cases or percentage of the total number of observations in each interval immediately gives an indication of the shape of the distribution of the population values, and is called a frequency distribution or histogram (Figure 4.1). The number of intervals is arbitrary, and the visual appearance of the distribution on a graph is affected by the number of intervals. Some information contained in the original dataset is lost when the observations are grouped and, in general, the fewer the number of intervals, the greater the loss. The number of intervals should ensure balance among accuracy, ease of communication, the use to which the information will be put, and the statistical tests to which the data will be subjected.

A third approach to organization is to form a cumulative frequency distribution, also called an ogive. A graph is constructed by plotting the cumulative number or percentage of observations against the ordered values of the element (Figure 4.2). The ogive representation of the data is useful for determining what proportion of the data is above or below a certain value. The proportion of values below a certain value, expressed as a percentage, is called a percentile; one

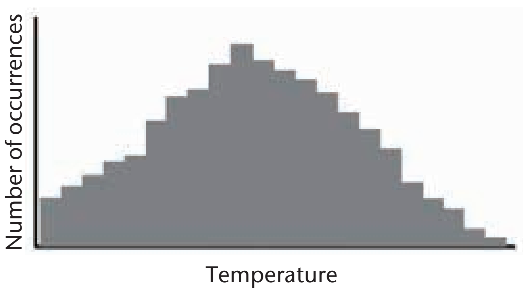



**Figure 4.1. Frequency distribution (histogram)**

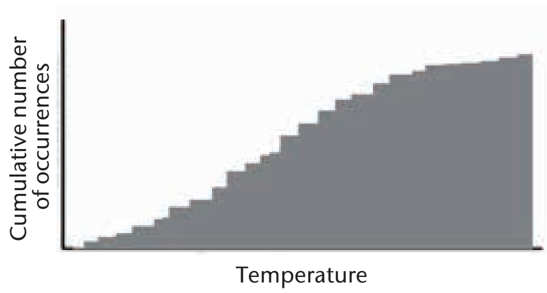



**Figure 4.2. Cumulative frequency distribution**

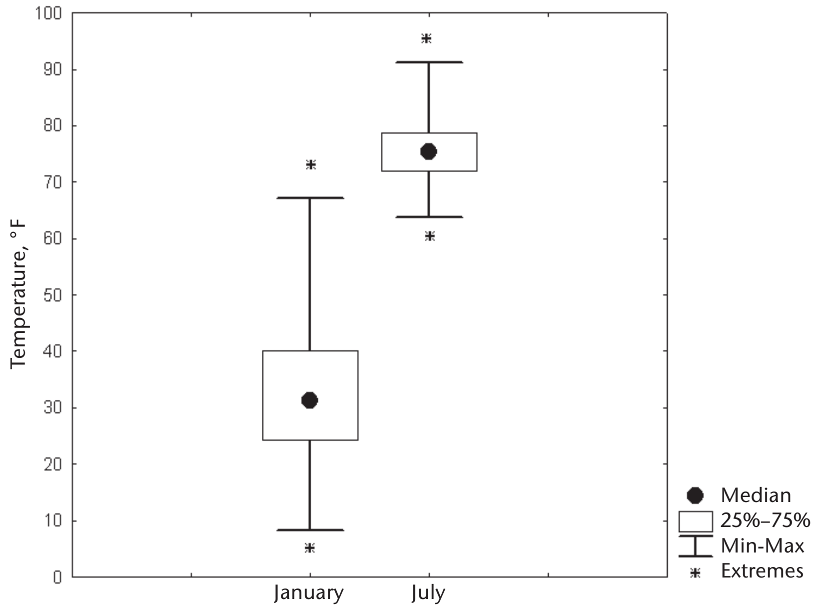



**Figure 4.3. Box plot**

per cent of the observations are smaller than the first percentile, two per cent are smaller than the second percentile, and so on. Similarly, a proportion based on tenths is called a decile; one tenth of the observations are below the first decile, two tenths are below the second decile, and so on. One based on quarters is called a quartile. One based on fifths is called a quintile and has particular use in calculations of precipitation normals (see 4.8.6).

Other visualizations include box plots (Figure 4.3), stem and leaf diagrams (Figure 4.4) and data arrays (Figure 4.5). If the sequence of the data is important, then a graph of observed values against time can be plotted to produce a time series (see 4.6). For data with two elements, such as wind speed and direction, scatter diagrams can be constructed by plotting the value of the first element against the value of the second (see 4.5.2). Wind roses also provide excellent depictions of wind information. Double-mass curves, which are frequently used by hydrologists and for data homogeneity detection, are constructed by plotting the cumulative value of one element against the cumulative value of the second element (see 5.2). Visualization techniques are limited only by the imagination of the analyst, but all techniques involve the sorting and classifying of the data. No matter which technique is used, the resulting graphic should be informative and should not inadvertently lead users to unsupported conclusions.

## 4.4 QUANTITATIVE SUMMARY DESCRIPTORS OF DATA

Rather than presenting the entire dataset to illustrate a particular feature, it is often useful to extract several quantitative summary measures. The summary measures help describe patterns of variation of observations. Understanding these patterns furthers the knowledge of the physical processes that underlie the observations, and improves inferences that can be made about past and current climate conditions.

Care must be taken to ensure that the contents of a dataset that are summarized by quantitative measures are really comparable. For example, a series of temperature observations may be comparable if they are all taken with the same instrumentation, at the same time each day, at the same location, and with the same procedures. If the procedures change, then artificial variations can be introduced in the dataset (see 3.3, 3.4 and 5.2). Sometimes the summary descriptors of a dataset identify unexpected variations; any unexpected patterns should be examined to determine whether they are artificially induced or real effects of the climate system.

### 4.4.1 **Data modelling of frequency distributions**

Data visualizations (see 4.3) provide a qualitative view of the structure of a series of observations. Shapes and patterns become apparent. Frequency distributions can be classified by their shape:

- Unimodal symmetrical curves: these curves are common for element averages, such as annual and longer-term average temperature. Generally, the longer the averaging period, the more symmetrical the distribution;
- Unimodal moderately asymmetrical curves: many curves of averaged data are mostly (but not quite) symmetrical;
- Unimodal strongly asymmetrical curves: these shapes are far from symmetrical and exhibit a high degree of skew; they are common for precipitation amounts and wind speeds;
- U-shaped curves: these curves are common for elements that have two-sided boundaries, such as the fraction of cloud cover (there are greater tendencies for skies to be mostly clear or mostly overcast).Multimodal or complex curves: these curves are common for elements observed daily in areas with strong seasonal contrasts. In such a case the

| Stem | | | | Leaf | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 5 | 7 | 8 | | | | | |
| 1 | 2 | 3 | 3 | 7 | 9 | | | |
| **2** | 1 | 3 | 4 | 4 | **5** | **5** | 7 | 8 |
| 3 | 2 | 2 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 4 | 5 | 5 | 6 | 6 | 6 | 6 | 7 |
| 5 | 2 | 3 | 3 | 3 | 4 | 8 | 8 | |
| 6 | 1 | 1 | 3 | 3 | 4 | 5 | | |
| 7 | 2 | 2 | 5 | 5 | 7 | | | |
| 8 | 4 | 6 | 7 | 7 | | | | |

**Figure 4.4. Example of a stem and leaf data display. The leading digit of the value of an observation is the stem, and the trailing digit is the leaf. In the table there are, for example, two observations of 25.**

| Year | | | | | | Month | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1961 | 23 | 33 | 44 | 50 | 59 | 64 | 79 | 76 | 61 | 50 | 44 | 32 |
| 1962 | 26 | 31 | 40 | 54 | 60 | 67 | 78 | 73 | 60 | 49 | 40 | 30 |
| 1963 | 27 | 35 | 43 | 55 | 58 | 68 | 77 | 72 | 58 | 52 | 43 | 32 |
| 1964 | 24 | 37 | 47 | 58 | 57 | 64 | 79 | 74 | 59 | 54 | 46 | 34 |
| 1965 | 27 | 32 | 43 | 56 | 57 | 65 | 76 | 74 | 58 | 53 | 47 | 44 |
| 1966 | 30 | 38 | 44 | 53 | 58 | 67 | 80 | 75 | 58 | 55 | 46 | 32 |
| 1967 | 19 | 35 | 47 | 55 | 61 | 66 | 74 | 73 | 60 | 56 | 43 | 30 |
| 1968 | 22 | 33 | 46 | 56 | 60 | 69 | 78 | 70 | 56 | 52 | 45 | 30 |
| 1969 | 28 | 37 | 43 | 51 | 56 | 70 | 76 | 72 | 54 | 52 | 44 | 34 |
| 1970 | 25 | 34 | 46 | 56 | 58 | 63 | 73 | 71 | 54 | 50 | 43 | 31 |

**Figure 4.5. Example of a data array**

frequency distribution curve built using the whole dataset may show a very characteristic bimodal shape. Datasets with very complex frequency distributions are likely to be better understood by stratifying the data a priori to reflect the different underlying processes.

More than one series of observations can be simplified by examining the distribution of observations of one variable when specified values of the other variables are observed. The results are known as conditional frequencies. The conditions are often based on prior knowledge of what can be expected or on information about the likelihood of certain events occurring. Conditional frequency analysis is especially useful in developing climate scenarios and in determining local impacts of events such as the El Niño–Southern Oscillation (ENSO) phenomenon and other teleconnection patterns (strong statistical relationships among weather patterns in different parts of the Earth).

One approach to summarizing the distribution of a set of observations is to fit a probability distribution to the observations. These distributions are functions with known mathematical properties that are characterized by a small number of parameters (typically no more than three). The functions are always constructed so that the relative magnitudes of different values reflect differences in the relative likelihoods of observing those values. Several common probability distributions, such as the normal (or Gaussian) distribution and the generalized extreme value distribution, describe situations that often occur in nature. If an observed frequency or conditional frequency distribution can be described by these known probability density functions, then the properties and relationships can be exploited to analyse the data and make probabilistic and statistical inferences. Some examples of probability density functions that may approximate observed frequency distributions of continuous data (where any value in a continuum may be observed) are shown in Figures 4.6 to 4.12.

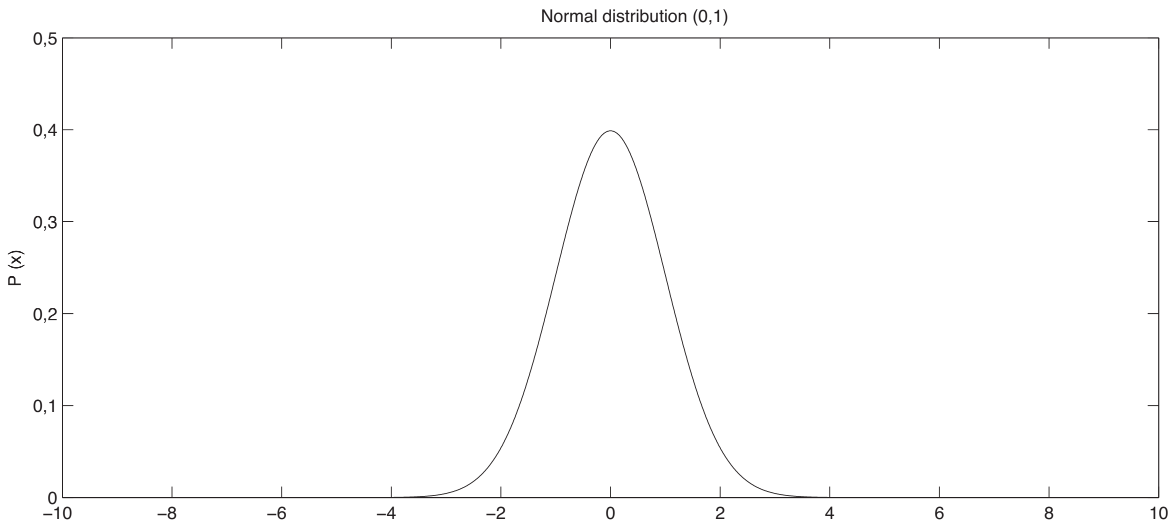



**Figure 4.6. Normal or Gaussian distribution: values over the range for the property being observed tend to cluster uniformly around a single value, such as annual average temperatures.**

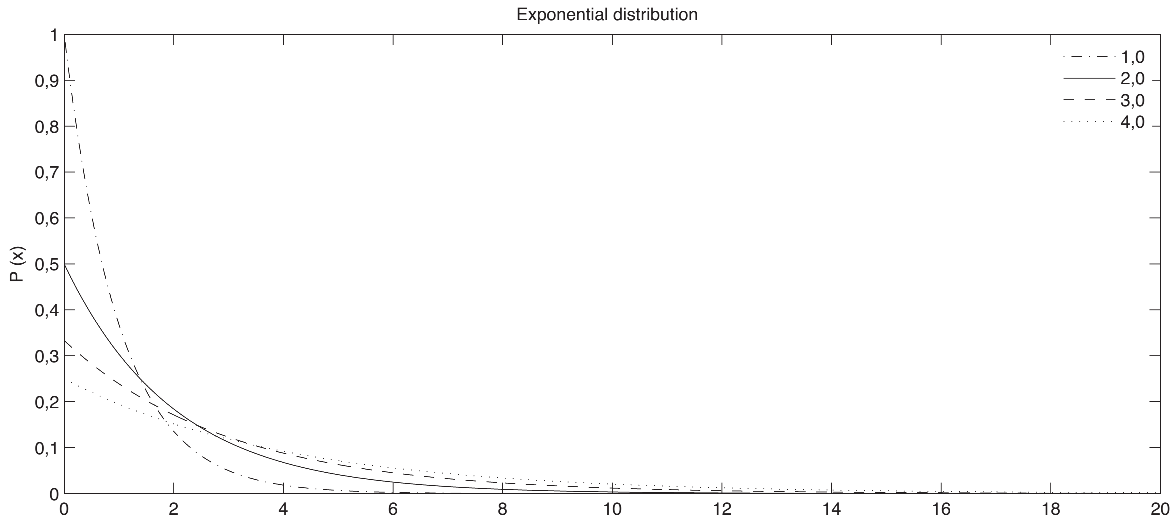



**Figure 4.7. Exponential distribution: describes the times between events in a process in which events occur continuously and independently at a constant average rate. It has been used in analyses of daily precipitation amounts.**

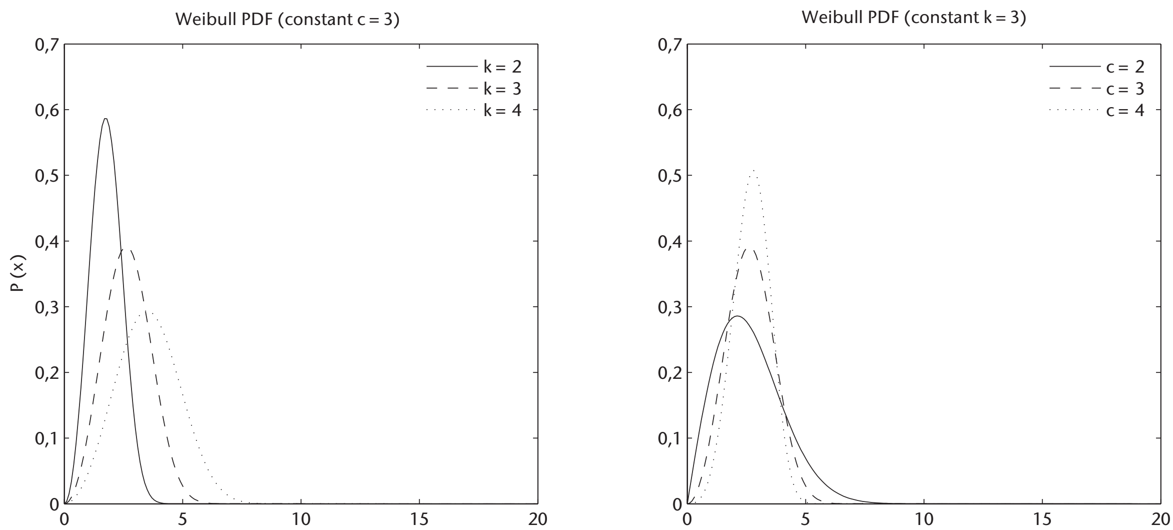



**Figure 4.8. Weibull distribution: describes the times between events in a process in which events occur continuously and independently at a varying rate. It has been used in analyses of wind speed.**

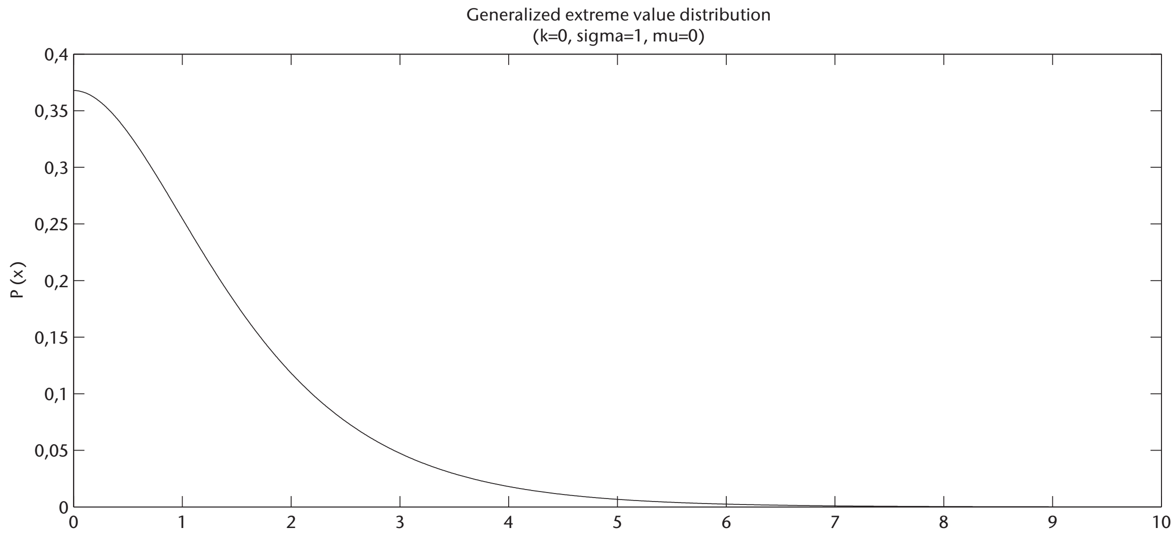



**Figure 4.9. Generalized extreme value distribution: used to model extreme values in a distribution.**

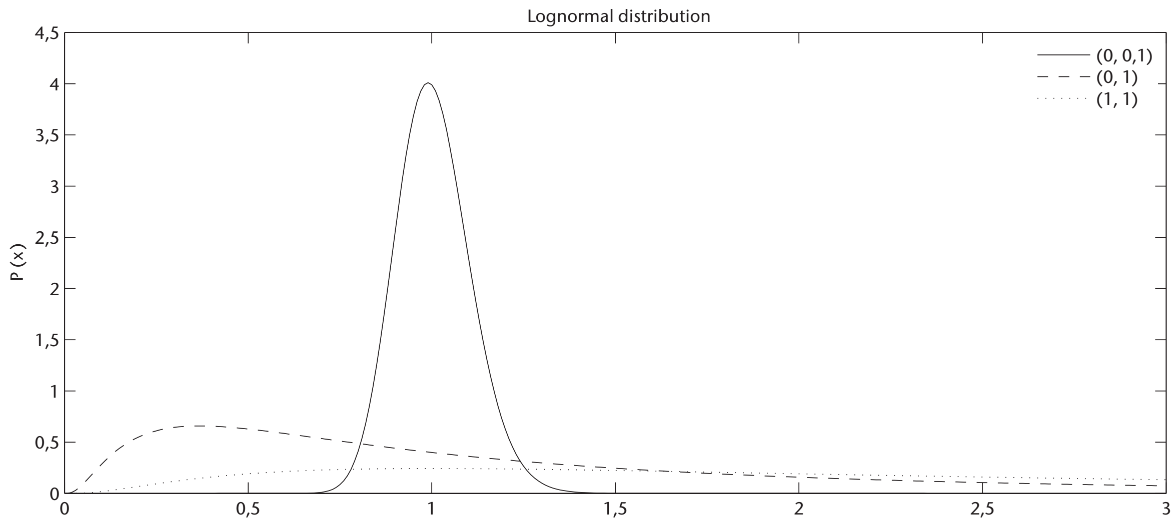



**Figure 4.10. Lognormal distribution: used when the logarithm of a distribution follows the normal distribution, such as that of air pollution particle concentrations.**

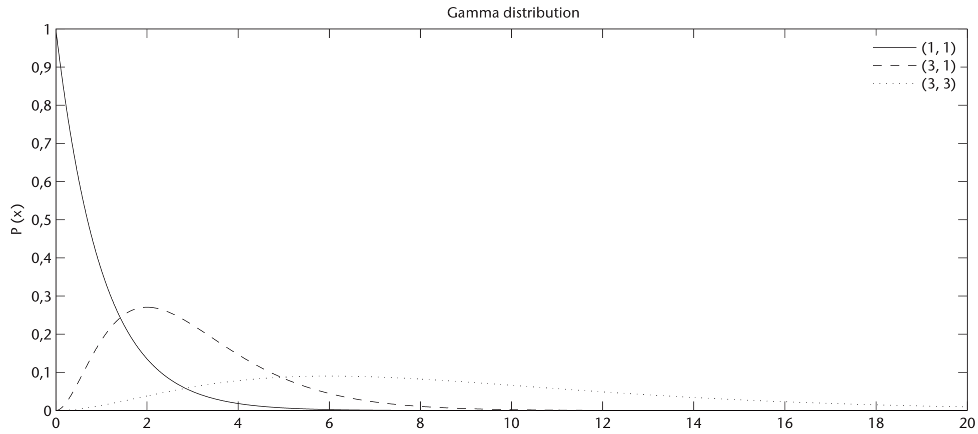



**Figure 4.11. Gamma distribution: describes distributions that are bounded at one end and skewed, such as precipitation data.**

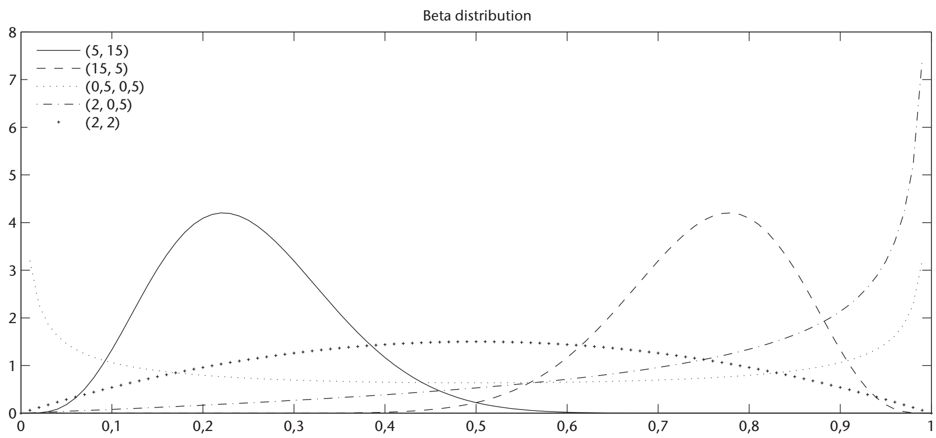



**Figure 4.12. Beta distribution: describes distributions that are bounded at both ends, such as cloud amounts.**

Statistical frequency distributions are also defined for describing data that can attain only specific, discrete values. An example is the number of days in which precipitation occurs in a month; only integer values are possible up to a maximum value of 31. Some of the statistical frequency distributions that describe discrete data are shown in Figures 4.13 and 4.14.

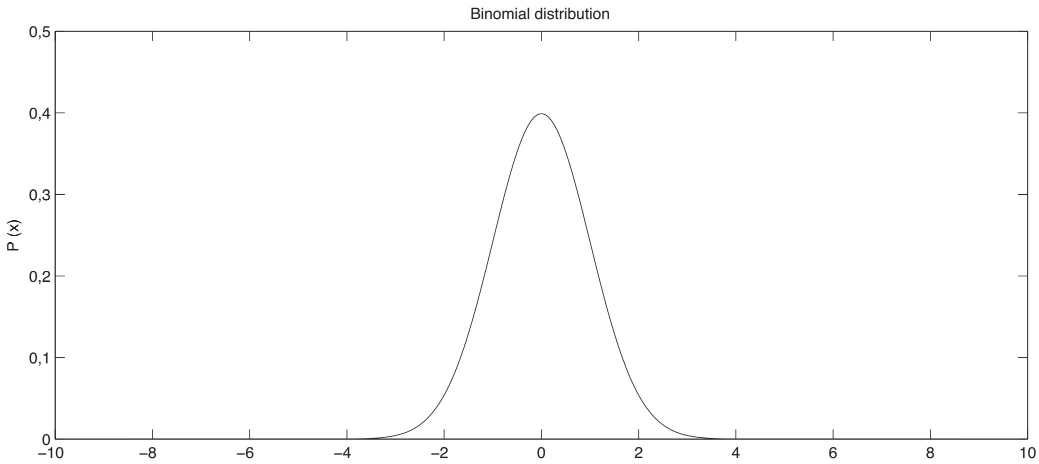



**Figure 4.13. Binomial distribution: describes two discrete outcomes such as the occurrence or non-occurrence of an event.**

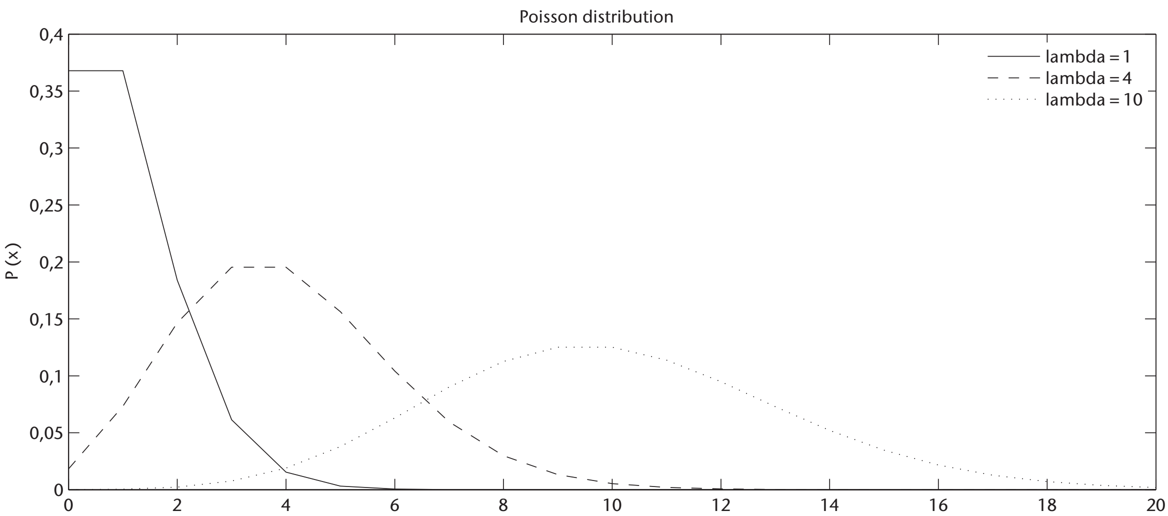



**Figure 4.14. Poisson distribution: describes rare events such as the frequency of occurrence of tropical storms.**

The visual, qualitative display of an observed frequency distribution often suggests which of the statistical frequency distributions may be appropriate to describe the data. If a statistical frequency distribution with known probabilistic relationships can be used to describe the data, then inferences about the data can be made. There are specific approaches to fitting distributions to observations, such as the methods of moments, probability-weighted moments (L-moments) and maximum likelihood. An understanding of the statistical theory underlying a distribution function is necessary for making proper inferences about the data to which it is being applied.

Any series of observations can be modelled by mathematical functions that reproduce the observations. One needs to take care in applying such techniques for a number of reasons. For example, the application of a mathematical function generally assumes that the observational dataset, usually a sample, fairly represents the population from which it is drawn, and that the data contain no errors (see 3.4). The main intent of fitting a function is to approximate the distribution of the observations. If the fit is acceptable, then with just a few parameters, the summary function of the observations should provide a realistic description of the data that is compatible with the underlying physics in a smoothed form that ignores data errors. A secondary intent is to describe the data within a theoretical framework that is sufficiently simple so that statistical inferences can be made. Overfitting with a mathematical model may lead to an unrealistic description of the data with too much weight placed on data errors or random factors that are extraneous to the processes being studied. The degree of smoothing is usually determined by how the data are to be used and by what questions the climatologist is trying to answer.

How well a summary function describes the observations can be determined by examining the differences between the observations and the values produced by the function. Objective goodness-of-fit tests should be applied. Datasets can usually be modelled by more than one function, and the test measures can be compared to find the best or most useful fit. The chi-square and Kolmogorov-Smirnov tests are commonly used. The chi-square goodness-of-fit test assumes that data values are discrete and independent (no observation is affected by any other observation). If the sum of the squares of the differences between observed and fitted frequencies exceeds a threshold that depends on the sample size, then the fit should be regarded as inappropriate. This test is sensitive to the number of intervals. The Kolmogorov-Smirnov test hypothesizes that if the maximum absolute difference between two continuous cumulative frequencies of independent observations is larger than a critical value, then the distributions are likely to be different. This test is effective if the dataset has a large number of observations.

### 4.4.2 **Measures of central tendency**

Observations often tend to cluster around a particular value. Measures of central tendency are designed to indicate a central value around which data tend to cluster. Measures of central tendency are no substitute for all the detailed information contained in the complete set of observations. Calculation of a single measure is often inadequate to describe the manner in which the data tend to concentrate because it does not consider variation of the observations. Any measure of central tendency should be accompanied by a measure of the degree of variation in the values of the observations from which the central tendency is derived.

The arithmetic mean, commonly known as average, is one of the most frequently used statistics in climatology. It is calculated by simply dividing the sum of the values by the number of values. For observations that tend to cluster around a central value, the mean represents a number towards which the average in a very long time series of observations or other large dataset would converge as the number of data values increases. The mean is not representative of the central tendency of strongly asymmetrical distributions.

A weighted mean is calculated by assigning different levels of importance to the individual observations so that, for example, more trustworthy or more representative observations can be given more influence in the calculation of a mean. Weights can be determined by many methods. A common example is distance weighting, where weights are inversely related to a measure of distance. For example, distance weighting is often used when estimating a mean value representative of a specific location from observations taken in a region surrounding that location. The weights are generally mathematical relations that may have no inherent relation to the physical processes that are measured, but whenever possible the choice of weighting methods should try to take into account physical considerations. Generally speaking, weighting methods give good results when both physical and statistical properties vary continuously and quite slowly over the studied space and time.

A mean has several advantages: it is a very convenient standard of reference for fluctuations in the observations (since the sum of the departures from a mean is zero); it is easily calculated; means for different non-overlapping subsets of the whole observational record can be combined; and the error of an estimate of a mean from a sample is smaller than other measures of central tendency (see *On the Statistical Analysis of Series of Observations* (WMO-No. 415), section 3.1, examples 26 and 27).

Means, however, have limitations. Any single value may prove misleading when used to describe a series of observations. Very similar means may be computed from datasets or distributions that are totally different in their internal structure. For example, the mean of a bimodal distribution of cloud cover may be the same as the mean of a unimodal distribution, but the interpretation of the two means would be very different. The mean is greatly affected by exceptional and unusual values; a few extreme observations may destroy the representative character of the mean. Observations that do not cluster towards a central value are not well represented by a mean (for example, cloud cover, which often tends to cluster at either 0 or 8 oktas). A mean, to be useful, must convey a valid meaning with respect to the actual conditions described by the dataset and not be merely the result of a mathematical calculation.

The median is the middle of a cumulative frequency distribution; half the data are above the median and the other half are below. It is calculated by ordering the data and selecting the middle value. If there are an odd number of values, the median is the middle value. For an even number of values the median is located between the two middle values, generally as the mean (or a weighted mean) of the two. If the two middle values are identical, then this value is chosen as the median.

Extreme variations have less of an influence on the median than on the mean because the median is a measure of position. Since the median is based on the number of observations, the magnitude of extreme observations does not influence the median. The median is especially useful when observations tend to cluster around the centre but some of the observations are also very high or very low. As with the mean, data that do not cluster towards a central value are not well represented by the median.

The mode is the value in the dataset that occurs most often. Like the median, it is a positional measure. It is affected neither by the value (as is the mean) nor by the position of other observations (as is the median). Modes from small samples or from samples that have more than one cluster of observations are unreliable estimates of central tendency. If multiple concentrations of the observations are really typical (a multimodal distribution), then the dataset may possibly be comprised of dissimilar factors, each of which has a different central value around which the observations tend to cluster.

For elements having a circular nature, such as wind direction, the concept of mean can be ambiguous. The modal value, such as prevailing wind direction, is often a more useful measure of central tendency for elements that are measured by direction.

A quantity that has a magnitude only is called a scalar. A quantity that associates a direction with its magnitude is called a vector. For example, wind velocity is a vector as it has both speed and direction. Mathematically, a vector can be transformed into independent components, and these components can then be averaged and combined into a resultant mean vector. For example, the wind can be expressed as a combination of two different scalars, eastward speed and northward speed, with westward and southward speeds, respectively, having negative values. The central tendency of the wind velocity is the resultant vector formed from the central tendencies of the eastward and northward speeds. A mathematical resultant vector calculated from data with opposing directions and equal speeds will have a magnitude of zero; this calculation may not be meaningful in the context of describing a climate. An alternative approach that may be more meaningful for climate descriptions is to calculate an average scalar direction, ignoring the speed but accounting for circularity (for example, wind directions of 355 and 5 degrees are separated by 10 degrees and not 350 degrees), and an average scalar magnitude ignoring direction. An alternative is to combine the resultant vector direction with the scalar average magnitude.

In a perfectly symmetrical frequency distribution with one mode, such as the Gaussian distribution, the values of the mean, median and mode will be exactly the same. If the frequency distribution is skewed to high values (skewed to the right), the mean will have the highest value, followed by the median and then the mode. This sequence is reversed if the frequency distribution is skewed to low values (skewed to the left). These relationships (Figure 4.15) and the features of the measures (Table 4.1) should be considered whenever a measure of central tendency is selected to represent a dataset.

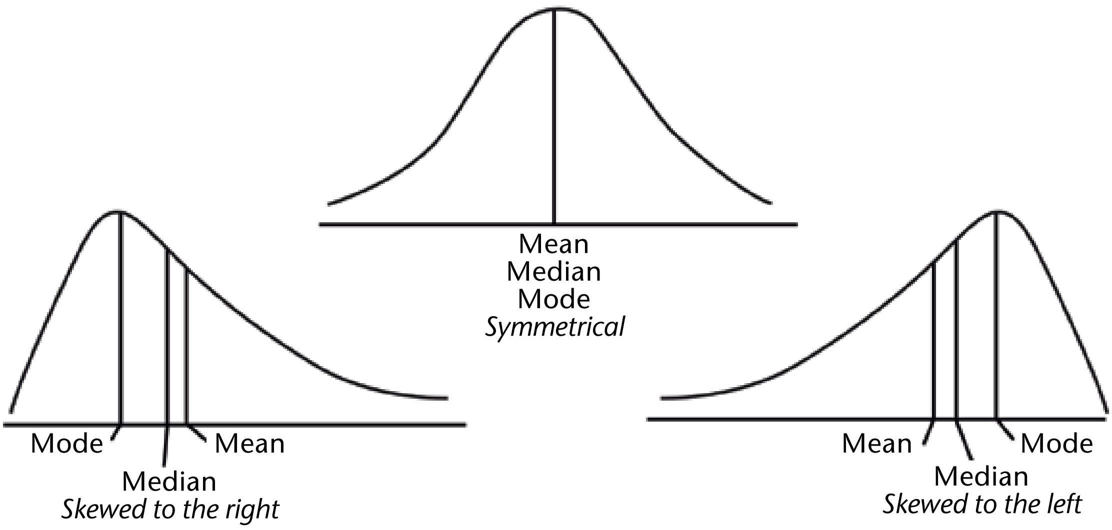



**Figure 4.15. Relationships among the mean, median and mode**

**Table 4.1. Comparison of some features of measures of central tendencies**

| *Feature* | *Mean* | *Median* | *Mode* |
|---|---|---|---|
| Affected by outliers | Yes | No | No |
| Representative of central tendency when frequency distributions are narrowly spread | Yes | Yes | Yes |
| Representative of central tendency when frequency distributions are widely spread | No | Maybe | No |
| Representative of central tendency when observations are clustered into more than one group | No | Maybe | No |
| Representative of central tendency when frequency distributions with one cluster are skewed | No | Yes | Yes |
| Ease of calculation | Easiest | Easy from ordered data | Easy from histogram |
| Departures sum to zero | Yes | Not always | Not always |
| Possibility for more than 1 | No | No | Yes |
| Indicator of variability | No | No | Only if more than one mode |

### 4.4.3 **Measures of variability**

Once a suitable estimate of the central tendency is chosen, it is possible to measure the variability of individual observations around that value. The measurement of variation and its explanation is of fundamental importance. However, a record of only a few observations generally gives a poor basis for judging the variability.

Variability can be measured absolutely or relatively. The deviation of each individual observation from the central tendency can be reduced to a value that represents and describes the entire dataset. This single number is the absolute variability.

The simplest measure of absolute variability is the range of the observations. The range is the difference between the highest and lowest values. Although easy to calculate, the range has many limitations. If the extreme values are very rare or they fall well beyond the bulk of observations, then the range will be misleading. The range provides no information about the nature of the frequency distribution within the extreme limits. The range also ignores the degree of concentration of the values almost entirely, and fails to characterize in a useful manner the dataset as a whole. Also, the range offers no basis for judging the reliability of the central tendency.

The interquartile range is another common measure of absolute variability. It is the difference between the third and first quartiles. The first quartile is the value of the ordered observations such that 25 per cent are below this value and 75 per cent are above, and the third quartile is the value of the ordered data such that 75 per cent are below this value and 25 per cent are above. The interquartile range is thus the range of the central 50 per cent of the ordered observations. When coupled with the median, it describes some of the characteristics of the frequency distribution. Other central ranges can be calculated in a similar way. The interdecile range, for example, is the difference between the 90th and 10th percentiles, and is the range of the middle 80 per cent of observations.

The average deviation is the mean of the absolute value of all the deviations of individual observations from the chosen measure of central tendency. While deviations may be calculated from the mean, median or mode, they should theoretically be calculated from the median because the sum of the absolute deviations from the median is less than or equal to the sum from either the mean or mode.

The standard deviation is the square root of the mean of the square of all the individual deviations from the mean. Deviations are taken from the mean instead of the median or mode because the sum of squares from the mean is a minimum. Squaring deviations gives greater weight to extreme variations. The standard deviation is used in the derivation of many statistical measures. It is also used extensively as a normative quantity to standardize different distributions for comparative purposes.

For comparative purposes, absolute measures of variability may have serious limitations. Comparisons should be made only if averages from which the deviations have been measured are approximately equal in value, and when the units of measurement are the same. For example, comparing standard deviations calculated from a temperature dataset and from a heating degree-day dataset is meaningless.

Often, comparisons are required when the means are not approximately equal or when the units of measurement are not the same. Some measure is therefore required that takes into account the mean from which deviations are measured, and that reduces different units of measurement to a common basis for the purposes of comparison. The relation of the absolute variability to the magnitude of the central tendency is the relative variability. One such measure is the coefficient of variation, which is the ratio of the standard deviation to the mean of a dataset.

### 4.4.4 **Measure of symmetry**

Skewness is a measure of the departure from symmetry. It is a relative and dimensionless measure, therefore allowing for comparisons among datasets. One simple measure of skewness is the difference between the mean and mode, divided by the standard deviation. Skewness is positive when the mean is greater than the mode and negative when the mode is greater than the mean. Other measures have also been defined, such as one based on the interquartile range and median.

Positive skewness is characteristic of some precipitation datasets that have a lower limit of zero but an unbounded upper limit. Daily maximum temperature datasets also often tend towards positive skewness, but daily minimum temperatures often tend towards negative skewness.

### 4.4.5 **Measure of peakedness**

Symmetrical frequency distributions may have different degrees of flatness in the central part of the distribution. Kurtosis is a dimensionless ratio that provides a relative measure for comparative purposes of the flatness or peakedness. Positive kurtosis indicates a narrow maximum in the centre of the frequency distribution, with frequencies falling sharply to low values away from the mean. Negative values indicate a large, flat central region, and are typical of many meteorological distributions, such as upper-air humidity.

### 4.4.6 **Indices**

The purpose of an index is to reduce complex conditions to a single number that retains some physical meaning and can be used to monitor a particular process. It expresses the relationship between observed and baseline conditions as a single value. The baseline is usually, but not always, the average climatic state. An example is the Palmer Drought Severity Index, which is a summary comparison of a complex water balance system of precipitation, evaporation, runoff, recharge and soil properties to climatically average conditions. Development of an index has four components: the selection of the elements that are to be included in the index; the selection and calculation of the baseline; the method of construction of the index; and the weights or importance of each of the included elements.

Examination and selection of the data to be included in the index often constitute a more complicated task than the actual calculation of the index. One of the concerns when choosing a baseline is that the characteristics of the observations used to define the baseline may change over time; it is essential that the observations used are homogeneous (see 5.2). Another concern is that the baseline should represent normal, standard or expected conditions, since most users of an index assume that the baseline represents such conditions. When selecting a baseline, care should be taken to explicitly define what is to be compared and for what purpose.

Selection of weights is a critical consideration. The importance of each element contributing to the index should be weighted in relation to the purpose of calculating the index. If the index is to be calculated in the future, care must also be taken to periodically review the contribution of each element for changes in, for example, importance, data accuracy, measurement and processing.

There are other indices developed specifically to assess the changing climate based on observations. These indices are referred to as climate change indices (see *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WCDMP-No. 72, WMO-TD No. 1500)).

## 4.5 **CORRELATION**

One often needs to detect or describe the relationship between or among elements. A relationship may be evident from visual displays of data, but quantitative measures are often calculated. Correlation is a measure that quantifies a relationship. No matter which measure is calculated, it is important to note that correlation does not imply a cause and effect relationship, but only that elements behave in a similar way. Often, factors other than those being investigated could be responsible for the observed association; many apparent relationships in meteorology and climatology are generally too complex to be explained by a single cause. Just as a positive or negative correlation does not imply causation, a zero correlation does not necessarily imply the absence of a causative relationship.

### 4.5.1 **Contingency tables**

Contingency tables are a simple yet effective way of discovering important relationships among factors, especially for large datasets. Contingency tables are most often formed from qualitative descriptors (such as mild, moderate or severe), or from dichotomous variables (an event did or did not occur). They can also be formed from the joint frequency of two elements, such as wind speed and direction or the diurnal distribution of visibility. Table 4.2 is an example of a contingency table.

**Table 4.2. Contingency table of highway accidents and visibility observations**

| | *Visibility below 200 meters* | *Visibility above 200 meters* | *Total* |
|---|---|---|---|
| Accident occurred | 16 | 4 | 20 |
| No accident occurred | 13 | 332 | 345 |
| Total | 29 | 336 | 365 |

The independence of the elements of a contingency table is often assessed by using a chi-square test. When this test is used, the serial dependence often found in climatological time series, according to which an observation is more likely to be similar to its preceding observation than dissimilar (see 4.6), violates the assumptions of the test, so that conclusions drawn from the test may be suspect.

### 4.5.2 Measures of correlation

A scatter diagram is another simple yet useful tool for visualizing relationships. It can show a relationship between two elements or the trend of one element over time, or whether any useful relationship exists at all. Figures 4.16 and 4.17 are examples of scatter diagrams. The association among elements and temporal patterns can sometimes be summarized by a correlation measure. The correlation coefficient is the most commonly used measure of association. Another measure, the Spearman rank correlation, is also sometimes used.

The correlation coefficient is a number between –1 and +1. It measures the linear relationship between two elements. A coefficient of zero implies no similarity of behaviour between elements. Figure 4.16 is an example of the pattern that is expected when two elements are very weakly correlated. A coefficient of +1 indicates that as the value of one element increases, the value of the other element also increases in direct proportion. Figure 4.17 is an example of the pattern that is expected when two elements have a strong positive correlation. A coefficient of –1 indicates that as the value of the first element increases, the value of the other element decreases in inverse proportion. One of the problems with using a simple correlation coefficient is that the implied relationship is linear. Often, meteorological elements are related in a non-linear manner, and the dataset may need to be transformed (see 5.4) prior to the calculation of a correlation coefficient.

The Spearman rank correlation measures agreement between the ordered ranks of two datasets. The measure is again a number between –1 and +1. If the observations in the two sets do not maintain the same relative order, then the measure will be low or negative; if they have a similar relative order, the coefficient will be high and positive. The Spearman measure is less sensitive to extremes than the correlation coefficient; it measures linear association, and sometimes indicates non-linear association.

## 4.6 TIME SERIES

An ordering of observations by time of occurrence is called a time series (Figure 4.18). A graph of data values plotted against time is an important qualitative tool for identifying time-related variations. In climatology, a trend is an interesting characteristic because it summarizes the historical behaviour of observations of an element. Linear trends are examined most often for a given time series, but a trend may sometimes be better described in non-linear terms, such as a curve, or even as an abrupt upward or downward shift. Trends, whether linear or non-linear, are generally sustained in climate series for a finite period, which may be quite long. Over time the climate system has often shown trends in one direction, which are eventually followed by a

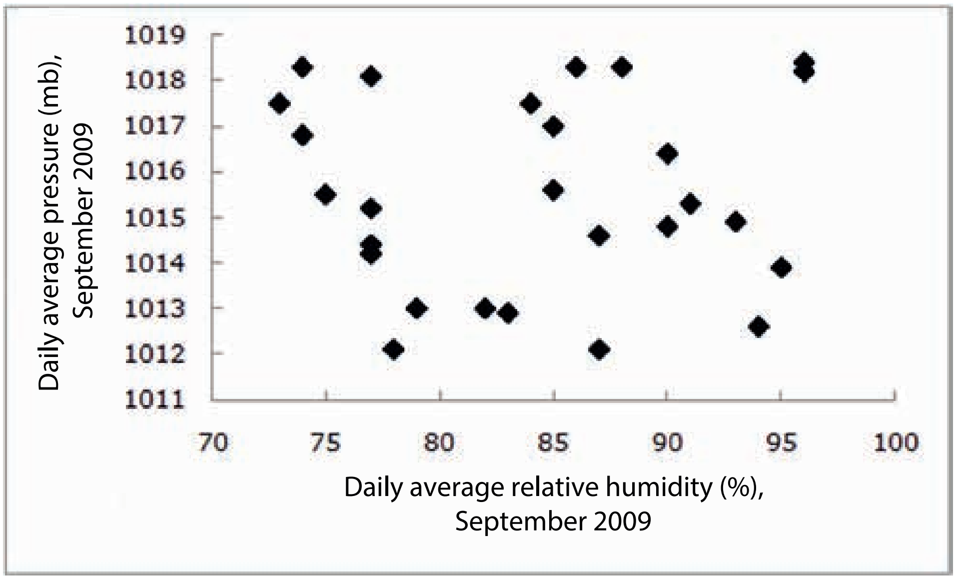



**Figure 4.16. Scatter diagram with weak correlation**

reversal. What might appear as a sustained trend in the most recent period of a climate record could be part of a slow oscillation related to multi-decadal variations that cannot be clearly seen because the time interval of the apparent sustained trend is only a part of the whole oscillation, or because the nature of the series projected into the future is unknown. Anthropogenic climate change poses a particularly difficult challenge in this regard since human decisions will likely play a part in determining, for example, how long the global warming trend observed over the past century will be sustained. The complete study of a time series requires the identification not only of the trends, but also of the periodic or quasi-periodic oscillations, and of the irregular or apparently random variations exhibited in the data. The goal of time series analysis is to understand how the variability in a time series is distributed as a function of the timescale.

Commonly in meteorology and climatology, successive observations tend to be more similar to each other than dissimilar. The measure used to summarize the relationship between each

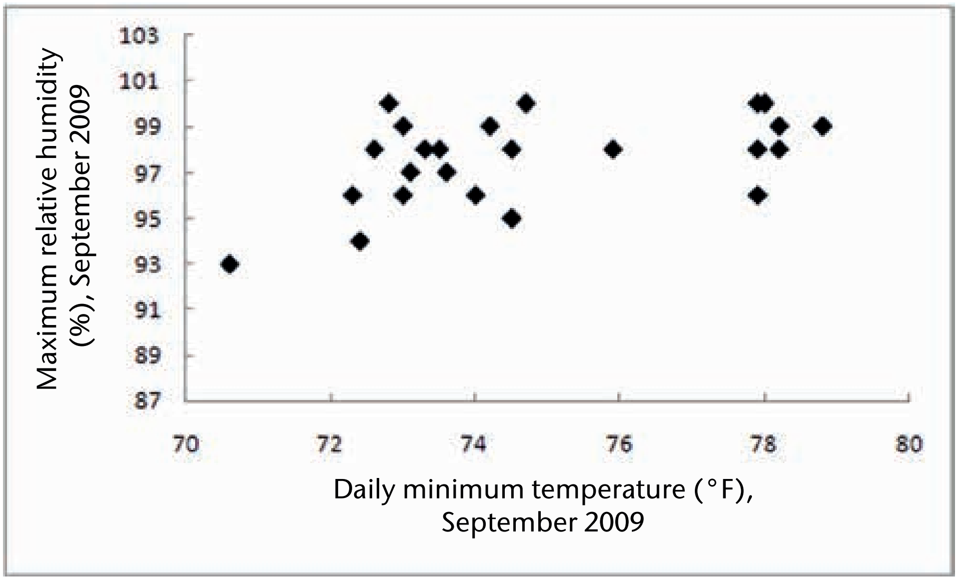



**Figure 4.17. Scatter diagram with strong positive correlation**

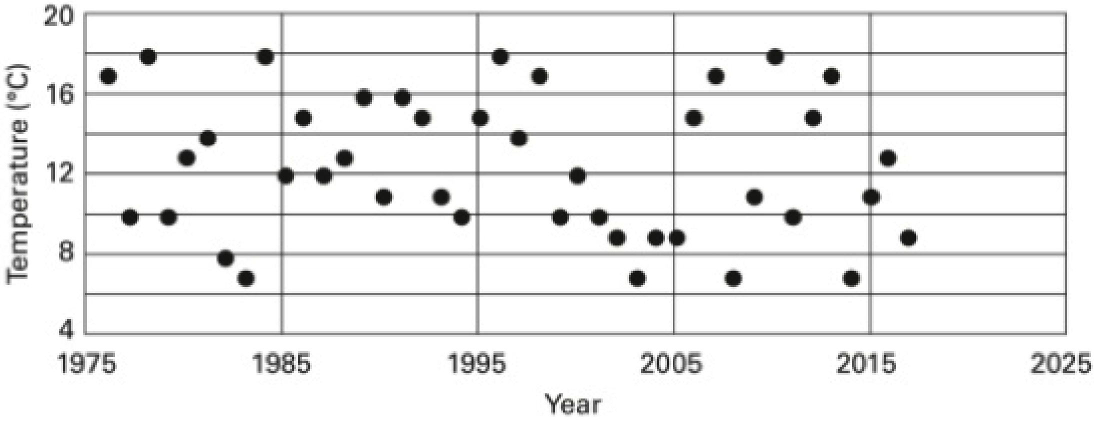



**Figure 4.18. Time series of March average temperatures**

observation and its predecessor is the autocorrelation coefficient. This measure is calculated in the same way as the correlation coefficient (see 4.5.2), with the exception that the second series is the same as the first, but shifted by one or more time steps.

Measures that summarize trends depend on the kind of trend being isolated. Linear trends are represented by the slope of a straight line. Non-linear trends are represented by the coefficients of the mathematical variables defining the equation for a curve fit, such as the coefficients of a polynomial function. Similarly, periodic features are also represented by the coefficients of the mathematical variables defining the oscillations, such as the frequency, phase and amplitude of trigonometric functions.

## 4.7 INTERPRETATION OF SUMMARY CHARACTERISTICS OF CLIMATE

Although it is possible to calculate numerous summary measures, it may not be appropriate to use them to describe the dataset. All measures that reduce observations with the purpose of detecting and describing a climate signal or relationship are based on assumptions, and if these assumptions are not valid the summary measures may be misleading. There are four issues that must be considered in detail before using summary measures: dataset errors, inhomogeneity, independence of observations and neglect of important factors.

Often, data are erroneous because of recording errors (such as the transposition of numbers), garbled communications, misunderstanding of coding practices by an observer, processing errors (for example, improper conversion from degrees Fahrenheit to degrees Celsius), computer program logic and coding errors, and incorrect identifier information (location or time) of a value. These types of error (see 3.3.3 and 3.4) do not relate to the physical conditions being observed, and they can contaminate data, with the result that improper conclusions are drawn from a data analysis.

Faulty inferences are often made when quantitative measures are used to compare data that are not really comparable, such as when comparing inhomogeneous observations. If possible, any dataset being analysed should be made homogeneous (see 5.2).

Many meteorological datasets violate the assumption of independence. Before a dataset is summarized, care should be taken to remove, if possible, dependence among observations. For example, the effect of known annual cycles can be largely removed by summarizing departures from the known cycle. Another example: if persistence (autocorrelation) is known to affect a series of observations, as sometimes occurs with daily temperatures observed during a synoptic surface high-pressure event, this should be taken into account by the analytical model. If dependence is not accounted for by the models, subsampling by selecting only one observation of the several available during the persistent event would remove the persistence affecting all the observations taken during the event. Care must also be taken in this process, however, so as not to mask any underlying oscillation, which can lead to an incorrect analysis.

An incomplete or erroneous explanation can result from presenting quantitative evidence concerning only one factor while ignoring other important factors. An example is comparing temperatures over a cold season at a coastal location and a continental location. The averages may be similar enough to suggest that the climates are the same, but such a conclusion would not be drawn if the greater variability at the continental location were taken into account.

Specific statistical assumptions concerning, for example, the consistency and homogeneity of the data or the nature of the dependence between observations, are implicit in all statistical analysis techniques. These assumptions should be clearly identified and assessed by the analyst, and the interpretation of a summary measure should be tempered by the extent to which the assumptions are satisfied. If any of the assumptions are violated, then the interpretation of a summary measure should be changed to account for the violations. The usual interpretation of the measure may still suffice, but the real or suspected violations should be disclosed with the measure. For example, if annual average temperatures are calculated from a dataset that is known from the metadata to be inhomogeneous, then the assumption that all data are comparable is violated. This violation, and its effect on the calculation of the mean, should be disclosed.

## 4.8 NORMALS

Climate normals are used for two principal purposes. They serve as a benchmark against which recent or current observations can be compared, including providing a basis for many anomaly-based climate datasets (for example, global mean temperatures). They are also widely used, implicitly or explicitly, to predict the conditions most likely to be experienced in a given location.

Historical practices regarding climate normals (as described in previous editions of this Guide, the *Technical Regulations* (WMO-No. 49), Volume I, Definitions, and the *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No. 1188), 1.2.3) date from the first half of the twentieth century. The general recommendation was to use 30-year periods of reference. The 30-year period of reference was set as a standard mainly because only 30 years of data were available for summarization when the recommendation was first made. The early intent of normals was to allow comparison among observations from around the world. The use of normals as predictors slowly gained momentum over the course of the twentieth century.

Traditionally, climatological normals have focused on the mean value of a climate element over a period of time. As discussed in 4.4.2, the mean is an incomplete description of the climate, and many applications require information about other aspects of that element's frequency distribution and statistical behaviour, such as the frequency of extended periods when a value is above a threshold. Extreme values of an element over a specified period, and other statistical descriptors of the frequency distribution of an element (such as the standard deviation of daily or monthly values), are useful descriptors of the climate at a location and should be included with datasets of normals.

Many NMHSs calculate daily normals along with monthly and annual normals. Although not required by WMO, daily normals illustrate the non-random pattern of daily variations of an element that cannot be captured with monthly normals. They are calculated by averaging the values of an element for a specified calendar date over a period of time. The observed values are usually smoothed by 3- to 7-day moving averages or binomial smoothing to reduce the effects of random high-frequency temporal variability of weather systems. Another smoothing approach is to fit the series of daily averages calculated from the observations with spline, trigonometric or polynomial smoothing functions; these smoothed series become the daily normals (see 5.8).

### 4.8.1 Period of calculation

There is a need for a relatively stable reference period for long-term climate variability assessments and climate change monitoring. Historically, climatological standard normals were calculated every 30 years for 30-year periods (1901–1930, 1931–1960, 1961–1990 and so on). For the specific purpose of long-term climate monitoring, the normals calculated for the period 1 January 1961–31 December 1990 are referred to as a stable WMO reference period, and should be retained in perpetuity or until such time as a compelling scientific reason to change the period arises.

However, there is also a need for more frequent calculations of climatic normals in a changing climate. For example, global temperature datasets have been calculated as anomalies from the reference period 1961–1990. Using a more recent averaging period, such as 1981–2010 (see Figure 4.19), results in an improvement in "predictive accuracy" for elements that show a secular trend (that is, where the time series shows a consistent rise or fall in its values when measured over a long term). Also, 1981–2010 normals would be viewed by many users as more "current" than 1961–1990. Climatological standard normals are calculated every 10 years for 30-year periods at the start of every decade from the year ending with digit 1 (1981–2010, 1991–2020, etc.). Recalculation of the normals every 10 years also requires recalculation of the numerous datasets that use the normals as a reference (degree days, departures from normal, etc.). The advent of modern computing and increasingly modern database systems should make these calculations relatively easy.

A number of studies have found that 30 years is not generally the optimal averaging period for normals used for future outlooks. The optimal period for temperatures is often substantially shorter than 30 years, but the optimal period for precipitation is often substantially greater than 30 years. *The Role of Climatological Normals in a Changing Climate* (WMO/TD-No. 1377) and other references at the end of this chapter provide much detail on the predictive use of normals of several elements. For predictive uses, NMHSs are encouraged to prepare averages and period averages. The optimal length of record for predictive use of normals varies with element, geography and secular trend. In general, the most recent 5- to 10-year period of record has as much predictive value as a 30-year record. Shorter periods allow normals to be calculated for a much wider range of stations than is usually possible for a standard normals period. For elements that show a substantial underlying trend (such as mean temperature), predictive accuracy is improved by updating the averages and period averages frequently.

In any publication of reference periods, normals and averages, and also in any publication that uses them for analysis and display of climate variability, it is important to document the period used for the calculation and the calculation methodologies.

### 4.8.2 Stations for which normals and averages are calculated

Climate normals and averages should be calculated for as wide a range of stations as possible, subject to the requirement that a station meet standards for the amount and completeness of available data. If possible, they should be calculated at least for all stations whose data are distributed on the Global Telecommunication System (see the *Manual on the Global Telecommunication System* (WMO-No.386), Attachment I-3, 2.1(c), and Attachment I-5, 4.1(e) and Table D, 3.1(a)).

### 4.8.3 Homogeneity of data

The data used in the calculation of climate normals and averages should, as far as possible, be homogeneous. The issue of homogeneity is addressed more fully in 5.2. In the context of climate normals and averages, homogeneity issues that require particular attention are changes of site location; changes of observation procedure, including changes of observation time; changes of instrument type; changes of instrument exposure over time; and changes in the processing of data.

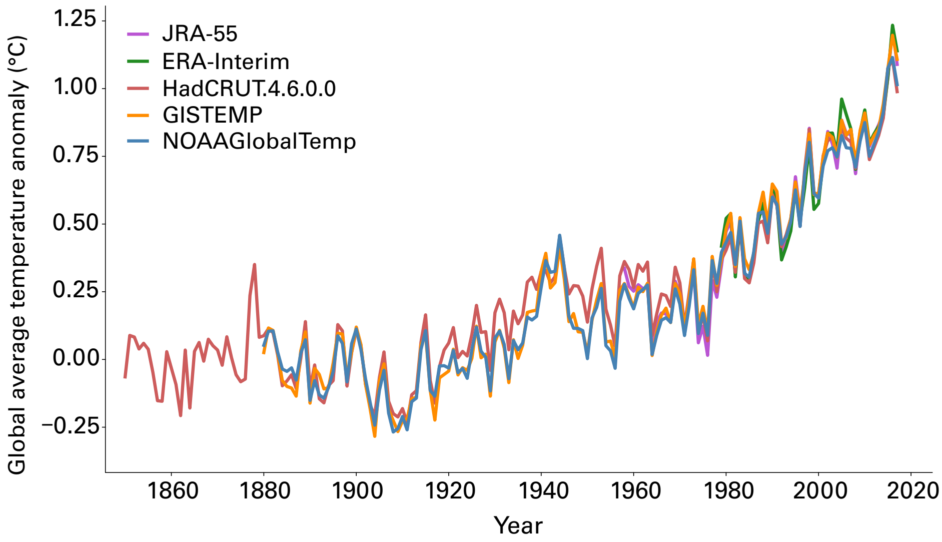



**Figure 4.19. Global mean temperature anomalies, with respect to the 1850–1900 baseline for the five global datasets (courtesy of the Met Office Hadley Centre, United Kingdom)**

In practice, at many locations, it will not be possible to construct a suitably homogenous dataset. It may instead be necessary to produce normals from a composite of two or more parts of an inhomogeneous record. An option is to make adjustments to the earlier part of a record to make it as homogeneous as possible with the most recent data.

### 4.8.4 **Missing data**

Normals calculated from incomplete datasets can be biased. For example, if one year in a period was particularly cold, a normal calculated without data from that year would be higher than a normal that did include that year. As there is often considerable autocorrelation in climatological data, consecutive missing observations can have a greater impact on normals than the same number of missing observations scattered randomly through the period in question.

As a guide, normals or period averages should be calculated only when values are available for at least 80 per cent of the years of record, with no more than three consecutive missing years. An alternative option, when there is an extended period of missing data but reasonably complete data after that time, is to calculate a period average using only data from the years following the break in the record.

Annual normals or averages should be calculated as the mean or sum (as appropriate) of the 12 monthly normals or averages, without consideration of the varying lengths of the months (see *WMO Guidelines on the Calculation of Climate Normals* (WMO-No.1203), 4.4). No missing monthly normals are permitted in the calculation of annual normals.

It is recommended that a monthly value should not be calculated if more than ten daily values or five or more consecutive daily values are missing. In the case of elements for which the monthly value is a sum of daily values rather than a mean (such as for rainfall or sunshine), a monthly value should be calculated only if either all daily observations are available, or if any missing days are incorporated in an observation accumulated over the period of missing data on the day when observations resume. The *Calculation of Monthly and Annual 30-Year Standard Normals* (WMO/TD-No. 341) recommends stricter criteria for calculating averages, with the limits being more than five missing days in total, or more than three consecutive missing days.

### 4.8.5 Average daily temperature

There are many methods for calculating an average daily temperature. These include methods that use a daily maximum and daily minimum, 24 hourly observations, synoptic observations and observations at certain specified hours in the course of a day. The best statistical approximation of an average is based on the integration of continuous observations over a period of time; the higher the frequency of observations, the more accurate the average. Practical considerations generally preclude the calculation of a daily average from a large number of observations evenly distributed over a 24-hour period because many observing sites do not measure an element continuously. For comparative purposes, a standard processing methodology is desirable for all stations worldwide, with as many stations as possible.

All ordinary climatological stations observe a daily maximum and minimum temperature (see 2.2.1). Hence, the recommended methodology for calculating average daily temperature is to take the mean of the daily maximum and minimum temperatures. Even though this method is not the best statistical approximation, its consistent use satisfies the comparative purpose of normals. An NMHS should also calculate daily averages using other methods if these calculations improve the understanding of the climate of the country.

### 4.8.6 Precipitation quintiles

Quintiles of precipitation are used to relate an observed monthly precipitation total to the frequency distribution of values observed over the period for which normals have been calculated. No universally accepted method exists for the calculation of quintile boundaries, and the choice of method can make a substantial difference to the resulting values. The recommended procedure for calculating the boundaries, however, is as follows:

(a) For any month, the 30 monthly values of precipitation recorded during the 30-year normal period are listed in ascending order. The list is then divided into five groups of quintiles of six values each. The first quintile contains the six lowest values for the month in question that have been observed during the 30-year period, the second quintile the next six lowest values, and so on to the fifth quintile, which contains the six highest values;

(b) The boundary between two adjacent quintiles is set halfway between the top value of one quintile and the first value of the next. The quintile index is the number of the lowest quintile containing the monthly precipitation in the month for which the report is being prepared, with the following special rules:

   (i) If the precipitation is 0, use index 0 if this has not occurred during the reference period, use 1 if it has occurred but fewer than 6 times, use 2 if it has occurred between 7 and 12 times, use 3 if it has occurred 13 to 18 times, and so on;

   (ii) If the precipitation is less than any value in the reference period: use index 0 (regardless of whether the precipitation is 0);

   (iii) If the precipitation is greater than any value in the reference period: use index 6.

## REFERENCES


World Meteorological Organization, 1989: *Calculation of Monthly and Annual 30-Year Standard Normal*s (WMO/TD-No. 341, WCDP-No. 10). Geneva.

———, 1990: *On the Statistical Analysis of Series of Observations* (WMO-No. 415). Geneva.

———, 2004: *Handbook on CLIMAT and CLIMAT TEMP Reporting* (WMO/TD-No. 1188). Geneva.

———, 2007: *The Role of Climatological Normals in a Changing Climate* (WMO/TD-No. 1377, CDMP-No. 61). Geneva.

———, 2009: *Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation* ( WCDMP-No. 72, WMO-TD No.1500). Geneva.

———, 2015: *Technical Regulations* (WMO-No. 49), Volume I : *General Meteorological Standards and Recommended Practices*. Geneva.

———, 2015: *Manual on the Global Telecommunication System* (WMO-No.386). Geneva.

———, 2017: *WMO Guidelines on the Calculation of Climate Normals* (WMO-No.1203). Geneva.


## FURTHER READING


Angel, J.R., W.R. Easterling and S.W. Kirtsch, 1993: Towards defining appropriate averaging periods for climate normals. *Climatological Bulletin*, 27:29–44.

Asnani, G.C., 1993: *Tropical Meteorology* (2 vols.). Pune, Sind Society.

Barnston, A.G. and R.E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115:1083–1126.

Bhalme, H.N. and D.A. Mooley, 1980: Large-scale droughts/floods and monsoon circulation. *Monthly Weather Review*, 108:1197–1211.

Bhalme, H.N., D.A. Mooley and S.K. Jadhav, 1984: Large-scale April pressure index connected with the southern oscillation and its potential for prediction of large-scale droughts over India. *Mausam*, 35(3):355–360.

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *Journal of Geophysical Research* 111, D12106.

Dixon, K.W. and M.D. Shulman, 1984: A statistical evaluation of the predictive abilities of climatic averages. *Journal of Climate and Applied Meteorology*, 23:1542–1552.

Guttman, N.B., 1989: Statistical descriptors of climate. *Bulletin of the American Meteorological Society*, 70:602–607.

Huang, J., H.M. van den Dool and A.G. Barnston, 1996: Long-lead seasonal temperature prediction using optimal climate normals. *Journal of Climate*, 9:809–817.

Lamb, P.J. and S.A. Changnon, 1981: On the “best” temperature and precipitation normals: the Illinois situation. *Journal of Applied Meteorology*, 20:1383–1390.

Matheron, G., 1973: The intrinsic random functions and their applications. *Advances in Applied Probability*, 5:207–221.

Rao, G.A., S.V. Datar and H.N. Srivastava, 1992: Study of drought indices in relation to rice crop production over some States of India. *Mausam*, 43(2):169–174.

Srivastava, A.K., P. Guhathakurta and S.R. Kshirsagar, 2003: Estimation of annual and seasonal temperatures over Indian stations using optimal normals. *Mausam*, 54:615–622.

Tukey, J.W., 1977: *Exploratory Data Analysis*. Reading, Massachusetts, Addison Wesley.

Von Storch, H. and F.W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge, Cambridge University Press.

Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. San Diego, Academic Press.

# CHAPTER 5. STATISTICAL METHODS FOR ANALYSING DATASETS

## 5.1 INTRODUCTION

This chapter introduces some of the statistical concepts and methods available to climatologists, but does not provide detailed specifics of complex subjects. Some statistical methods are given only cursory treatment, while others are ignored. The references at the end of the chapter and textbooks on statistical theory and methods provide more detailed information. Two references that should be on every climatologist's bookshelf are *Some Methods of Climatological Analysis* (WMO-No. 199) and *On the Statistical Analysis of Series of Observations* (WMO-No. 415). Since new and improved statistical and analytical methodologies are rapidly emerging, climatologists should keep abreast of current techniques that have practical applications in climatology.

The main interest in the use of observed meteorological or climatological data is not to describe the data (see Chapter 4), but to make inferences from a limited representation (the observed sample of data) of complex physical events that are helpful to users of climatological information. The interpretation of climatological data usually involves both spatial and temporal comparisons of characteristics of frequency distributions. These comparisons answer common questions such as:

- Are average temperatures taken over a specific time interval at different locations the same?
- Is the variability of precipitation the same at different locations?
- Is the diurnal temperature range at a location changing over time, and if so, how?
- What is the likelihood of occurrence of tropical storms in an area?

Inferences are based directly on probability theory, and the use of statistical methods to make inferences is therefore based on formal mathematical reasoning. Statistics can be defined as the pure and applied science of creating, developing and applying techniques such that the uncertainty of inductive inferences may be evaluated. Statistics is the tool used to bridge the gap between the raw data and useful information, and it is used for analysing data and climate models and for climate prediction. Statistical methods allow a statement of the confidence of any decision based on application of the procedures.

The confidence that can be placed in a decision is important because of the risks that might be associated with making a wrong decision. Observed data represent only a single sample of the physical system of climate and weather, and are generally observed with some level of error. Conclusions can be correct or incorrect. Quantitative factors that describe the confidence of the decisions are therefore necessary to properly use the information contained in a dataset.

## 5.2 HOMOGENIZATION

Analysis of climate data to detect changes and trends is more reliable when homogenized datasets are used. A homogeneous climate dataset is one in which all the fluctuations contained in its time series reflect the actual variability and change of the represented climate element. Most statistical methods assume that the data under examination are as free from instrumentation, coding, processing and other non-meteorological or non-climatological errors as possible. Meteorological or climatological data, however, are generally neither homogeneous nor free from error. Errors range from systematic (they affect a whole set of observations the same way, such as constant instrument calibration errors or improper conversion of units) to random (any one observation is subject to an error that is as likely to be positive as negative, such as parallax differences among observers reading a mercury barometer).

The best way to keep the record homogeneous is to avoid changes in the collection, handling, transmission and processing of the data. It is highly advisable to maintain observing practices and instruments as unchanged as possible (see *Guide to the GCOS Surface and Upper-air Networks: GSN and GUAN* (WMO-TD No. 1106), 3.4. Unfortunately, most long-term climatological datasets have been affected by a number of factors not related to the broader-scale climate. These include changes in geographical location; local land use and land cover; instrument types, exposure, mounting and sheltering; observing practices; calculations, codes and units; and historical and political events. Some changes may cause sharp discontinuities such as steps (for example, following a change in instrument or site), while others may cause gradual biases (resulting, for example, from increasing urbanization in the vicinity of a site). In both cases, the related time series become inhomogeneous, and these inhomogeneities may affect the proper assessment of climatic trends. Note that site changes do not always affect observations of all elements, nor do changes affect observations of all elements equally. The desirability of a homogeneous record stems primarily from the need to distil and identify changes in the broader-scale climate. There are some studies, however, that may require certain "inhomogeneities" to be reflected in the data, such as an investigation of the effects of urbanization on local climate or of the effects of vegetation growth on the microclimate of an ecosystem.

Statistical tests should be used in conjunction with metadata in the investigation of homogeneity. In cases in which station history is documented well and sufficient parallel measurements have been conducted for relocations and changes of instrumentation, a homogenization based on this qualitative and quantitative information should be undertaken. Therefore, the archiving of all historical metadata is of critical importance for an effective homogenization of climatological time series and should be a special concern for all meteorological services (see Chapters 2 and 3).

After the metadata analysis, statistical tests may find additional inhomogeneities. The tests usually depend on the timescale of the data; the tests used for daily data are different from those used for monthly data or other timescales. The results of such statistical homogenization procedures have to be checked again with the existing metadata. In principle, any statistical test that compares a statistical parameter of two data samples may be used. But usually, special homogeneity tests that check the whole length of a time series in one run are used. Both non-parametric tests (in which no assumptions about statistical distributions are made) and parametric tests (in which frequency distribution is known or correctly assumed) can be used effectively.

When choosing a homogeneity test, it is very important to keep in mind the shape of the frequency distribution of the data. Some datasets have a bell-shaped (normal or Gaussian) distribution; for these a parametric approach works well. Others (such as precipitation data from a site with marked inter-annual variability) are not bell-shaped, and rank-based non-parametric tests may be better. Effects of serial autocorrelation, the number of potential change points in a series (documented with metadata and undocumented), trends and oscillations, and short periods of record that may be anomalous should also be considered when assessing the confidence that can be placed in the results from any test.

Many approaches rely on comparing the data to be homogenized (the candidate series) with a reference series. A reference time series ideally has to have experienced all of the broad climatic influences of the candidate, but none of its possible and artificial biases. If the candidate is homogeneous, when the candidate and reference series are compared by differencing (in the case of elements measured on an interval scale, such as temperature) or by calculating ratios or log ratios (for elements measured on a proportional scale, such as precipitation), the resulting time series will show neither sudden changes nor trends, but will oscillate around a constant value. If there are one or more inhomogeneities, however, they will be evident in the difference or ratio time series. An example of an observed candidate series and a reference series is shown in Figure 5.1, and an example of a difference series revealing an inhomogeneity in a candidate series is shown in Figure 5.2.

Reference time series work well when the dataset has a large enough number of values to ensure a good climatological relation between each candidate and the neighbouring locations used in building the reference series, and when there are no inhomogeneities that affect all or most

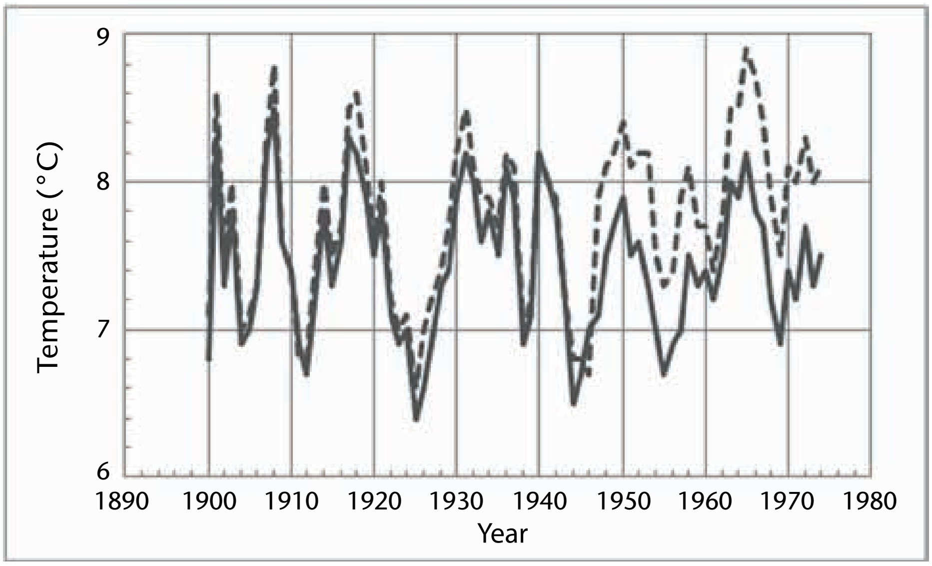



**Figure 5.1. Example of a candidate time series (dashed line) and a reference time series (solid line)**

of the stations or values available. In general, a denser network is needed for climatic elements or climatic types with a high degree of spatial variability (for examples, more data points are needed for precipitation than for temperature, and more data points are needed to homogenize temperature in a highly variable temperature climate than in a less variable one). When a change in instruments occurs at about the same time in an entire network, the reference series would not be effective because all the data points would be similarly affected. When a suitable reference series cannot be constructed, possible breakpoints and correction factors need to be evaluated without using any data from neighbouring stations.

The double-mass graph is often used in the field of hydrometeorology for the verification of measures of precipitation and runoff, but it can be used for most elements. The accumulated total from the candidate series is plotted against the accumulated total from the reference series for each available period. If the ratio between the candidate and reference series remains constant over time, the resultant double-mass curve should have constant slope. Any important variation in the slope or the shape of the curve indicates a change in the relationship between the

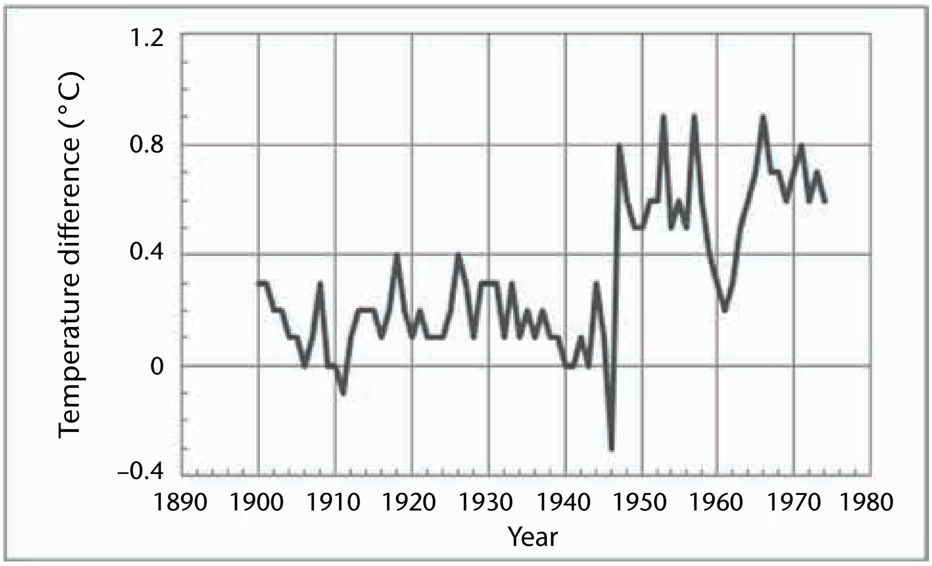



**Figure 5.2. Example of a difference time series**

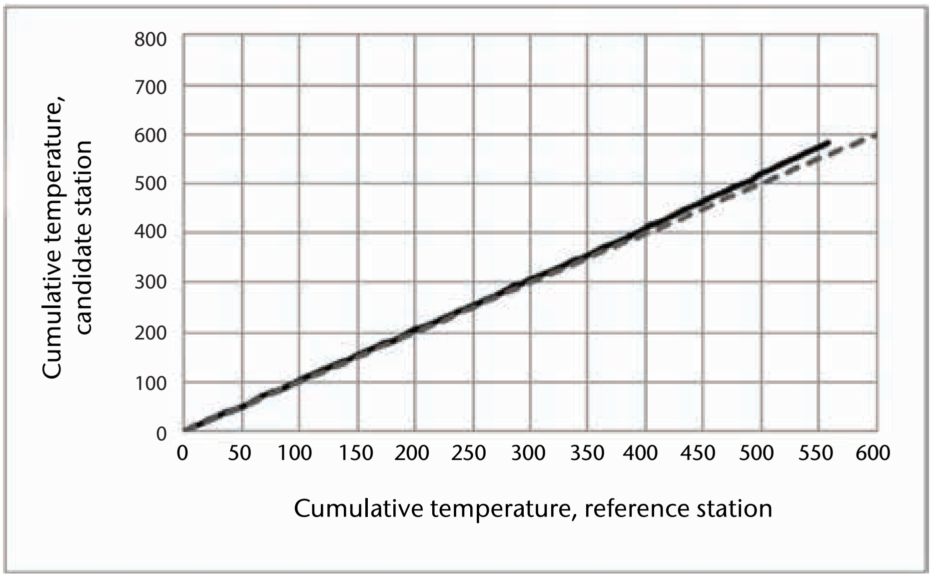



**Figure 5.3. Example of a double-mass graph with the dashed line representing a slope of 1**

two series. Since variations may occur naturally, it is recommended that the apparent changes of the slope be shown for a well-defined continuous period lasting at least five years and that they be consistent with events referenced in the metadata records of the station before concluding inhomogeneity. Figure 5.3 shows a double-mass graph for the same data used in Figures 5.1 and 5.2. Because it is often difficult to determine where on a double-mass graph the slope changes, a residual graph of the cumulative differences between the candidate and reference station data is usually plotted against time (Figure 5.4). The residual graph more clearly shows the slope change. The double-mass graph can be used to detect more than one change in proportionality over time. When the double-mass graph reveals a change in the slope, it is possible to derive correction factors by computing the ratio of the slopes before and after a change point.

There are several tests of stationarity (the hypothesis that the characteristics of a time series do not change over time). One is the runs test, which hypothesizes that trends and other forms of persistence in a sequence of observations occur only by chance. It is based on the total number of runs of directional changes in consecutive values. Too small a number of runs indicates

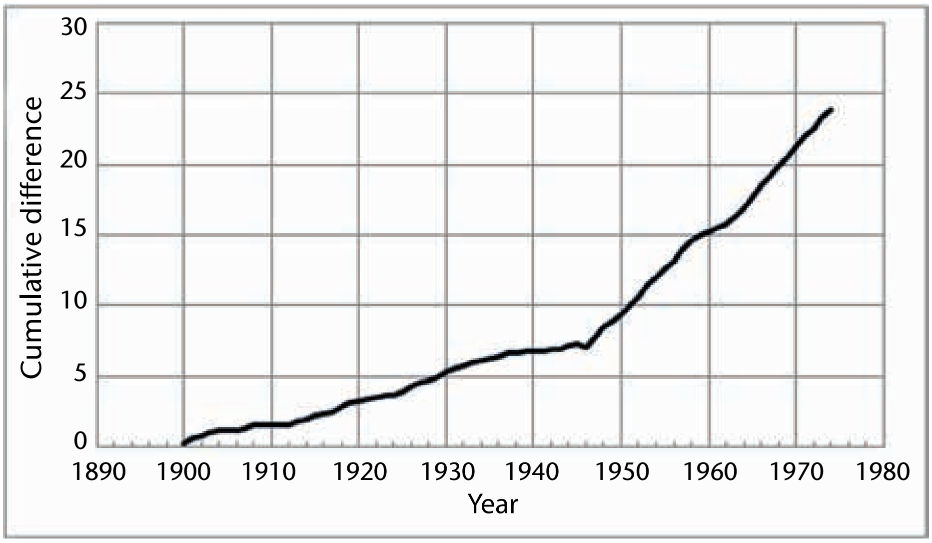



**Figure 5.4. Example of a residual double-mass graph**

persistence or trends, and too large a number indicates oscillations. Stationarity of central tendencies and variability between parts of a series are important. Techniques for examining these characteristics include both parametric and non-parametric methods.

Caution is needed when data are in sub-monthly resolution (such as daily or hourly observations) because one of the uses of homogeneous daily data is assessing changes in extremes. Extremes, no matter how they are defined, are rare events that are often created by a unique set of weather conditions. If few extreme data points are available for the assessment, determining the proper homogeneity adjustment for these unique conditions can be difficult. Extremes should be considered as part of the whole dataset, and they should therefore be homogenized not separately but along with all the data. Homogenization techniques for monthly, seasonal or yearly temperature data are generally satisfactory, but homogenization of daily data and extremes remains a challenge.

Although many objective techniques exist for detecting and adjusting the data for inhomogeneities, the actual application of these techniques remains subjective. At the very least, the decision about whether to apply a given technique is subjective. This means that independent attempts at homogenization may easily result in quite different data. It is important to keep detailed and complete documentation of each of the steps and decisions made during the process. The adjusted data should not be considered absolutely "correct", nor should the original data always be considered "wrong". The original data should always be preserved.

Homogeneity assessment and data adjustment techniques are an area of active development, and both the theory and practical tools are continuing to evolve. Efforts should be made to keep abreast of the latest techniques.

### 5.2.1 **Evaluation of homogenized data**

Evaluation of the results of homogeneity detection and adjustment is time-consuming but unavoidable, no matter which approach has been used. It is very important to understand which adjustment factors have been applied to improve the reliability of the time series and to make measurements comparable throughout their entire extent. Sometimes, one might need to apply a technique that has been designed for another set of circumstances (such as another climate, meteorological or climatological element, or network density), and it is important to analyse how well the homogenization has performed. For example, most techniques used to homogenize monthly or annual precipitation data have been designed and tested in rainy climates with precipitation throughout the year, and may have serious shortcomings when applied to data from climates with very dry seasons.

To assess corrections, one might compare the adjusted and unadjusted data to independent information, such as data from neighbouring countries, gridded datasets, or proxy records such as those from phenology, observation journals, or ice freeze and thaw dates. When using such strategies, one has also to be aware of their limitations. For example, gridded datasets might be affected by changes in the number of stations across time, or at a particular grid point they might not be well correlated with the original data from a co-located or nearby station.

Another approach is to examine countrywide, area-averaged time series for adjusted and unadjusted data and to see if the homogenization procedure has modified the trends expected from knowledge of the station network. For example, when there has been a widespread change from afternoon observations to morning observations, the unadjusted temperature data have a cooling bias in the time series, as the morning observations are typically lower than those in the afternoon.

If the homogenization results are valid, the newly adjusted time series as a whole will describe the temporal variations of the analysed element better than the original data. However, some single values may remain incorrect or made even worse by the homogenization. More complete descriptions of several widely used tests are available in the *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186), 3.4, Table 4, and in several of the references listed at the end of this chapter.

## 5.3 MODEL FITTING TO ASSESS DATA DISTRIBUTIONS

After a dataset is adjusted for known errors and inhomogeneities, the observed frequency distributions should be modelled by the statistical distributions described in 4.4.1 so that statistical methods can be exploited. A theoretical frequency distribution can be fitted to the data by inserting estimates of the parameters of the distribution, where the estimates are calculated from the sample of observed data. The estimates can be based on different amounts of information or data. The number of unrelated bits of information or data that are used to estimate the parameters of a distribution is called degree of freedom. Generally, the higher the number of degrees of freedom, the better the estimate will be. When the smooth theoretically derived curve is plotted with the data, the degree of agreement between the curve fit and the data can be visually assessed.

Examination of residuals is a powerful tool for understanding the data and suggests what changes to a model or data need to be made. A residual is the difference between an observed value and the corresponding model value. A residual is not synonymous with an anomalous value. An anomalous value is a strange, unusual or unique value in the original data series. A graphical presentation of residuals is useful for identifying patterns. If residual patterns such as oscillations, clusters and trends are noticed, then the model used is usually not a good fit to the data. Outliers (a few residual values that are very different from the majority of the values) are indicators of potentially suspicious or erroneous data values. They are usually identified as extremes in later analyses. If no patterns exist and if the values of the residuals appear to be randomly scattered, then the model may be accepted as a good fit to the data.

If an observed frequency distribution is to be fitted by a statistical model, the assumptions of the model and fitting process must be valid. Most models assume that the data are independent (one observation is unaffected by any other observation). Most comparative methods used in goodness-of-fit tests assume that errors are randomly and independently distributed. If the assumptions are not valid, then any conclusions drawn from such an analysis may be incorrect.

Once the data have been fitted by an acceptable statistical frequency distribution, meeting any necessary independence, randomness or other sampling criteria, and the fit has been validated (see 4.4), the model can be used as a representation of the data. Inferences can be made that are supported by mathematical theory. The model provides estimates of central tendency, variability and higher-order properties of the distribution (such as skewness or kurtosis). The confidence that these sample estimates represent real physical conditions can also be determined. Other characteristics, such as the probability of an observation's exceeding a given value, can also be estimated by applying both probability and statistical theory to the modelled frequency distribution. All of these tasks are much harder, if not impossible, when using the original data rather than the fitted frequency distribution.

## 5.4 DATA TRANSFORMATION

The normal or Gaussian frequency distribution is widely used, as it has been studied extensively in statistics. If the data do not fit the normal distribution well, applying a transformation to the data may result in a frequency distribution that is nearly normal, allowing the theory underlying the normal distribution to form the basis for many inferential uses. Transforming of data must be done with care so that the transformed data still represent the same physical processes as the original data and that sound conclusions can be made.

There are several ways to tell whether a distribution of an element is substantially non-normal. A visual inspection of histograms, scatter plots, or probability–probability (P–P) or quantile–quantile (Q–Q) plots is relatively easy to perform. A more objective assessment can range from simple examination of skewness and kurtosis (see 4.4.4 and 4.4.5) to inferential tests of normality.

Prior to applying any transformation, an analyst must ensure that the non-normality is caused by a valid reason. Invalid reasons for non-normality include mistakes in data entry and missing data values not declared missing. Another invalid reason for non-normality may be the presence of outliers, as they may well be a realistic part of a normal distribution.

The most common data transformations used for improving normality are the square root, cube root, logarithmic and inverse transformations. The square root makes values less than 1 relatively greater, and values greater than 1 relatively smaller. If the values can be positive or negative, a constant offset must be added before taking the square root so that all values are greater than or equal to 0. The cube root has a similar effect to the square root, but does not require the use of an offset to handle negative values. Logarithmic transformations compress the range of values, by making small values relatively larger and large values relatively smaller. A constant offset must first be added if values equal to 0 or lower are present. An inverse makes very small numbers very large and very large numbers very small; values of 0 must be avoided.

These transformations have been described in the relative order of power, from weakest to strongest. A good guideline is to use the minimum amount of transformation necessary to improve normality. If a meteorological or climatological element has an inherent highly non-normal frequency distribution, such as the U-shape distribution of cloudiness and sunshine, there are no simple transformations allowing the normalization of the data.

The transformations all compress the right side of a distribution more than the left side; they reduce higher values more than lower values. Thus, they are effective on positively skewed distributions such as precipitation and wind speed. If a distribution is negatively skewed, it must be reflected (values are multiplied by –1, then a constant is added to make all values greater than 0) to reverse the distribution prior to applying a transformation, and then reflected again to restore the original order of the element.

Data transformations offer many benefits, but they should be used appropriately, in an informed manner. All of the transformations described above attempt to improve normality by reducing the relative spacing of data on the right side of the distribution more than the spacing on the left side. However, the very act of altering the relative distances between data points, which is how these transformations aim to improve normality, raises issues in the interpretation of the data. All data points remain in the same relative order as they were prior to transformation, which allows interpretation of results in terms of the increasing value of the element. However, the transformed distributions will likely become more complex to interpret physically due to the curvilinear nature of the transformations. The analyst must therefore be careful when interpreting results based on transformed data.

## 5.5 TIME SERIES ANALYSIS

The principles guiding model-fitting (see 5.3) also guide time series analysis. A model is fitted to the data series; the model might be linear, curvilinear, exponential, periodic or some other mathematical formulation. The best fit (the fit that minimizes the differences between the data series and the model) is generally accomplished by using least-squares techniques (minimizing the sum of squared departures of the data from the curve fit). Residuals from the best fit are examined for patterns and, if any are found, the model is adjusted to incorporate the patterns.

Time series in climatology have been analysed with a variety of techniques that decompose a series either into time domain or into frequency domain components. A critical assumption of these models is that of stationarity (when characteristics of the series such as mean and variance do not change over the length of the series). This condition is generally not met by climatological data even if the data are homogeneous (see 5.2).

Gabor and wavelet analyses are extensions of the classical techniques of spectral analysis. Allowing subintervals of a time series to be modelled with different scales or resolutions relaxes the condition of stationarity. These analyses are particularly good at representing time series with subintervals that have differing characteristics. Wavelet analysis gives good results when the time series has spikes or sharp discontinuities. Compared to the classical techniques, it is particularly efficient for signals in which both the amplitude and frequency vary with time. One of the main advantages of these "local" analyses is the ability to present time series of climate processes in the coordinates of frequency and time, studying and visualizing the evolution of various modes of variability over a long period. They are used not only as a tool for identifying non-stationary scales of variations, but also as a data analysis tool to gain an initial understanding of a dataset. There have been many applications of these methods in climatology, such as in studies of the El Nino–Southern Oscillation (ENSO) phenomenon, the North Atlantic Oscillation, atmospheric turbulence, space–time precipitation relationships and ocean wave characteristics.

These methods, however, do have some limitations. The most important limitation of wavelet analysis is that an infinite number of wavelet functions are available as a basis for an analysis, and results often differ depending on which wavelet is used. This makes interpretation of results somewhat difficult because different conclusions can be drawn from the same dataset if different mathematical functions are used. It is therefore important to relate the wavelet function to the physical world prior to selecting a specific wavelet. Gabor and wavelet analysis techniques are emerging fields and, although the mathematics has been defined, future refinements in techniques and application methodology may mitigate the limitations.

Other common techniques for analysing time series are autoregression and moving average analyses. Autoregression is a linear regression of a value in a time series against one or more prior values in the series (autocorrelation). A moving average process expresses an observed series as a function of a random series. A combination of these two methods is called a mixed autoregressive and moving average (ARMA) model. An ARMA model that allows for non-stationarity is called a mixed autoregressive integrated moving average (ARIMA) model. These regression-based models can be made more complex than necessary, resulting in overfitting. Overfitting can lead to the modelling of a series of values with minimal differences between the model and the data values, but since the data values are only a sample representation of a physical process, a slight lack of fit may be desirable in order to represent the true process. Other problems include non-stationarity of the parameters used to define a model, non-random residuals (indicating an inappropriate model), and periodicity inherent in the data but not modelled. Split validation is effective in detecting model overfitting. Split validation refers to the practice of developing a model based on a portion of the available data and then validating the model on the remaining data that were not used in the model development.

Once the time series data have been modelled by an acceptable curve, and the fit validated, the mathematical properties of the model curve can be used to make assessments that would not be possible using the original data. These include measuring trends, cyclical behaviour, or autocorrelation and persistence, together with estimates of the confidence of these measures.

## 5.6 MULTIVARIATE ANALYSIS

Multivariate datasets are a compilation of observations of more than one element over time and/or space. These datasets are often studied for many different purposes. The most important purposes are to see if there are simpler ways of representing a complex dataset, if observations fall into groups that can be classified, and if interdependence exists among elements. Such datasets are also used to test hypotheses about the data. The sequence of the observations is generally not a consideration in multivariate analyses. Time series of more than one element are usually considered as a separate analysis topic with techniques such as cross-spectral analysis.

Principal components analysis, sometimes referred to as empirical orthogonal functions analysis, is a technique for reducing the dimensions of multivariate data. The process simplifies a complex dataset and has been used extensively in the analysis of climatological data. Principal components analysis methods decompose a number of correlated observations into a new set of uncorrelated (orthogonal) functions that contain the original variance of the data. These empirical orthogonal functions, also called principal components, are ordered so that the first component is the one explaining most of the variance, the second component explains the second-largest share of the variance, and so on. Since most of the variance is usually explained by just a few components, the methods are effective in reducing "noise" from an observed field. Individual components can often be related to a single meteorological or climatological element. The method has been used to analyse a diversity of fields that include sea-surface temperatures, regional land temperature and precipitation patterns, tree-ring chronologies, sea-level pressure, air pollutants, radiative properties of the atmosphere, and climate scenarios. Principal components have also been used as a climate reconstruction tool, such as in estimating a spatial grid of a climatic element from proxy data when actual observations of the element are not available.

Factor analysis reduces a dataset from a larger set of observations to a smaller set of factors. It is similar to principal components analysis except that the factors are not uncorrelated. Since a factor may represent observations of more than one element, meteorological or climatological interpretation of a factor is often difficult. The method has been used mainly in synoptic climatology studies.

Cluster analysis attempts to separate observations into groups with similar characteristics. There are many methods for clustering, and different methods are used to detect different patterns of points. Most of the methods, however, rely on the extent to which the distance between means of two groups is greater than the mean distance within a group. The measure of distance does not need to be the usual Euclidean distance, but it should obey certain criteria. One such criterion should be that the measure of distance from point A to point B is equal to the distance from point B to point A (symmetry). A second criterion is that the distance should be a positive value (non-negativity). A third criterion is that for three points forming a triangle, the length of one side should be less than or equal to the sum of the lengths of the other two sides (triangle inequality). A fourth criterion should be that if the distance from A to B is zero, then A and B are the same (definiteness). Most techniques iteratively separate the data into more and more clusters, thereby presenting the problem for the analyst of determining when the number of clusters is sufficient. Unfortunately, there are no objective rules for making this decision. The analyst should therefore use prior knowledge and experience in deciding when a meteorologically or climatologically appropriate number of clusters has been obtained. Cluster analysis has been used for diverse purposes, such as constructing homogeneous regions of precipitation, analysing synoptic climatologies, and predicting air quality in an urban environment.

Canonical correlation analysis seeks to determine the interdependence between two groups of elements. The method finds the linear combination of the distribution of the first element that produces the correlation with the second distribution. This linear combination is extracted from the dataset and the process is repeated with the residual data, with the constraint that the second linear combination is not correlated with the first combination. The process is again repeated until a linear combination is no longer significant. This analysis is used, for example, in making predictions from teleconnections, in statistical downscaling (see 6.3.5), in determining homogeneous regions for flood forecasting in an ungauged basin, and in reconstructing spatial wind patterns from pressure fields. A similar analysis that may be explored is singular value decomposition. This method, as with canonical correlation analysis, finds linear combinations of the two groups of elements such that the linear combinations attempt to capture the maximum possible covariance.

These methods all have assumptions and limitations. The interpretation of the results is very much dependent on the assumptions being met and on the experience of the analyst. Other methods, such as multiple regression and covariance analysis, are even more restrictive for most meteorological or climatological data. Multivariate analysis is complex, with numerous possible outcomes, and requires care in its application.

## 5.7 COMPARATIVE ANALYSIS

By fitting a model function to the data, be it a frequency distribution or a time series, it is possible to use the characteristics of that model for further analysis. The properties of the model characteristics are generally well studied, allowing a range of conclusions to be drawn. If the characteristics are not well studied, bootstrapping may be useful. Bootstrapping is the estimation of model characteristics from multiple random samples drawn from the original observational series. It is an alternative to making inferences from parameter-based assumptions when the assumptions are in doubt, when parametric inference is impossible, or when parametric inference requires very complicated formulas. Bootstrapping is simple to apply, but it may conceal its own set of assumptions that would be more formally stated in other approaches.

In particular, there are many tests available for comparing the characteristics of two models to determine how much confidence can be placed in claims that the two sets of modelled data share underlying characteristics. When comparing two models, the first step is to decide which characteristics are to be compared. These could include the mean, median, variance or probability of an event from a distribution, or the phase or frequency from a time series. In principle, any computable characteristic of the fitted models can be compared, although there should be some meaningful reason (based on physical arguments) to do so.

The next step is to formulate the null hypothesis. This is the hypothesis considered to be true before any testing is done, and in this case it is usually that the modelled characteristics are the same. The alternative hypothesis is the obverse, that the modelled characteristics are not the same.

A suitable test to compare the characteristics of the two models is then selected. Some of these tests are parametric, depending on assumptions about the distribution, such as normality. Parametric tests include the Student's t-test (for comparing means) and the Fisher's F-test (for comparing variability). Other tests are non-parametric, so they do not make assumptions about the distribution. They include sign tests (for comparing medians) and the Kolmogorov-Smirnov test for comparing distributions. Parametric tests are generally better (in terms of confidence in the conclusions), but only if the required assumptions about the distribution are valid.

The seriousness of rejecting a true hypothesis (or accepting a false one) is expressed as a level of confidence or probability. The selected test will show whether the null hypothesis can be accepted at the level of confidence required. Some of the tests will reveal at what level of confidence the null hypothesis can be accepted. If the null hypothesis is rejected, the alternative hypothesis must be accepted. Using this process, the analyst might be able to make the claim, for example, that the means of two sets of observations are equal with a 99 per cent level of confidence; accordingly, there is only a 1 per cent chance that the means are not the same.

Regardless of which hypothesis is accepted, the null or the alternative, the conclusion may be erroneous. When the null hypothesis is rejected but it is actually true, a Type I error has been made. When the null hypothesis is accepted and it is actually false, a Type II error has been made. Unfortunately, reducing the risk of a Type I error increases the risk of making a Type II error, so that a balance between the two types is necessary. This balance should be based on the seriousness of making either type of error. In any case, the confidence of the conclusion can be calculated in terms of probability and should be reported with the conclusion.

## 5.8 SMOOTHING

Smoothing methods provide a bridge between making no assumptions based on a formal structure of observed data (the non-parametric approach) and making very strong assumptions (the parametric approach). Making a weak assumption that the true distribution of the data can be represented by a smooth curve allows underlying patterns in the data to be revealed to the analyst. Smoothing increases signals of climatic patterns while reducing noise induced by random fluctuations. The applications of smoothing include exploratory data analysis, model building, goodness-of-fit of a representative (smooth) curve to the data, parametric estimation, and modification of standard methodology.

Kernel density estimation is one method of smoothing; examples include moving averages, Gaussian smoothing and binomial smoothing. Kernel smoothers estimate the value at a point by combining the observed values in the neighbourhood of that point. The method of combination is often a weighted mean, with weights dependent on the distance from the point in question.

The size of the neighbourhood used is called the bandwidth; the larger the bandwidth, the greater the smoothing. Kernel estimators are simple, but they have drawbacks. Kernel estimation can be biased when the region of definition of the data is bounded, such as near the beginning or end of a time series. As one bandwidth is used for the entire curve, a constant level of smoothing is applied. Also, the estimation tends to flatten peaks and valleys in the distribution of the data. Improvements to kernel estimation include correcting the boundary biases by using special kernels only near the boundaries, and by varying the bandwidths in different sections of the data distribution. Data transformations (see 5.4) may also improve the estimation.

Spline estimators fit a frequency distribution piecewise over subintervals of the distribution with polynomials of varying degree. Again, the number and placement of the subintervals affect the degree of smoothing. Estimation near the boundaries of the data is also problematic. Outliers can severely affect a spline fit, especially in regions with few observations.

A range of more sophisticated, often non-parametric, smoothers is also available. These include local maximum likelihood estimation, which is particularly useful when prior knowledge of the behaviour of the dataset can lead to a good "first guess" of the type of curve that should be fitted. These estimators are sometimes difficult to interpret theoretically.

With multivariate data, smoothing is more complex because of the number of possibilities of smoothing and the number of smoothing parameters that need to be set. As the number of data elements increases, smoothing becomes progressively more difficult. Most graphs are limited to only two dimensions, so visual inspection of the smoother is limited. Kernel density can be used to smooth multivariate data, but the problems of boundary estimation and fixed bandwidths can be even more challenging than with univariate data.

Large empty regions in a multivariate space usually exist unless the number of data values is very large. Collapsing the data to a smaller number of dimensions with, for example, principal components analysis, is a smoothing technique. The dimension reduction should have the goal of preserving any interesting structure or signal in the lower-dimension data while removing uninteresting attributes or noise.

One of the most widely used smoothing tools is regression. Regression models, both linear and non-linear, are powerful for modelling a target element as a function of a set of predictors, allowing for a description of relationships and the construction of tests of the strength of the relationships. These models are susceptible, however, to the same problems as any other parametric model in that the assumptions made affect the validity of inferences and predictions.

Regression models also suffer from boundary problems and unrealistic smoothing in subintervals of the data range. These problems can be solved by weighting subintervals of the data domain with varying bandwidths and by applying polynomial estimation near the boundaries. Regression estimates, which are based on least-squares estimation, can be affected by observations with unusual response values (outliers). If a data value is far from the majority of the values, the smooth curve will tend to be drawn closer to the aberrant value than may be justified. When using adjusted non-parametric smoothing, it is often difficult to unambiguously identify a value as an outlier because the intent is to smooth all the observations. Outliers could be a valid meteorological or climatological response, or they could be aberrant; additional investigation of the outlier is necessary to ensure the validity of the value. Regression estimates are also affected by correlation. Estimates are based on the assumption that all errors are statistically independent of each other; correlation can affect the asymptotic properties of the estimators and the behaviour of the bandwidths determined from the data.

## 5.9 ESTIMATING DATA

One of the main applications of statistics to climatology is the estimation of values of elements when few or no observed data are available or when expected data are missing. In many cases, the planning and execution of user projects cannot be delayed until there are enough meteorological or climatological observations; estimation is used to extend a dataset.

Estimation has also a role in quality control by allowing an observed value to be compared to its neighbours in both time and space. Techniques for estimating data are essentially applications of statistics, but should also rely on the physical properties of the system being considered. In all cases, it is essential that values statistically estimated be realistic and consistent with physical considerations.

Interpolation uses data that are available both before and after a missing value (time interpolation), or surrounding the missing value (space interpolation), to estimate the missing value. In some cases, the estimation of a missing value can be performed by a simple process, such as by computing the average of the values observed on both sides of the gap. Complex estimation methods are also used, taking into account correlations with other elements. These methods include weighted averages, spline functions, linear regressions and kriging. They may rely solely on the observations of an element, or take into account other information such as topography or numerical model output. Spline functions can be used when the spatial variations are regular. Linear regression allows the inclusion of many kinds of information. Kriging is a geostatistical method that requires an estimation of the covariances of the studied field. Cokriging introduces into kriging equations the information given by another independent element.

Extrapolation extends the range of available data values. There are more possibilities for error in extrapolated values because relations are used outside the domain of the values from which the relationships were derived. Even if empirical relations found for a given place or period of time seem reasonable, care must be taken when applying them to another place or time because the underlying physics at one place and time may not be the same at another place and time. The same methods used for interpolation can be used for extrapolation.

### 5.9.1 Mathematical estimation methods

Mathematical methods involve the use of only geometric or polynomial characteristics of a set of point values to create a continuous surface. Inverse distance weighting and curve fitting methods, such as spline functions, are examples. The methods are exact interpolators; observed values are retained at sites where they are measured.

Inverse distance weighting is based on the distance between the location for which a value is to be interpolated and the locations of observations. Unlike the simple nearest neighbour method (where the observation from the nearest location is chosen), inverse distance weighting combines observations from a number of neighbouring locations. Weights are given to the observations depending on their distance from the target location; close stations have a larger weight than those farther away. A "cut-off" criterion is often used, either to limit the distance from observation locations or the number of observations considered. Often, inverse squared distance weighting is used to provide even more weight to the closest locations. With this method no physical reasoning is used; it is assumed that the closer an observation location is to the location where the data are being estimated, the better the estimation. This assumption should be carefully validated since there may be no inherent meteorological or climatological reason to justify the assumption.

Spline fits suffer from the same limitation as inverse distance weighting. The field resulting from a spline fit assumes that the physical processes can be represented by the mathematical spline; there is rarely any inherent justification for this assumption. Both methods work best on smooth surfaces, so they may not result in adequate representations on surfaces that have marked fluctuations.

### 5.9.2 Estimation based on physical relationships

The physical consistency that exists among different elements can be used for estimation. For instance, if some global radiation measurements are missing and need to be estimated, elements such as sunshine duration and cloudiness could be used to estimate a missing value. Proxy data may also be used as supporting information for estimation. When simultaneous values at two stations close to each other are compared, sometimes either the difference or the quotient of the values is approximately constant. This is more often true for summarized data (for months or years) than for those over shorter time intervals (such as daily data). The constant difference or ratio can be used to estimate data. When using these methods, the series being compared should be sufficiently correlated for the comparison to be meaningful. Then, the choice of the method should depend on the time structure of the two series. The difference method can be used when the variations of the meteorological or climatological element are relatively similar from one station to the other. The ratio method can be applied when the time variations of the two series are not similar, but nevertheless proportional (this is usually the case when a series has a lower bound of zero, as with precipitation or wind speed, for example). In the event that those conditions are not met, particularly when the variances of the series at the two stations are not similar, these techniques should not be used. More complex physically consistent tools include regression, discriminant analysis (for the occurrence of phenomena) and principal components analysis.

Deterministic methods are based upon a known relation between an in situ data value (predictand) and values of other elements (predictors). This relation is often based on empirical knowledge about the predictand and the predictor. The empirical relation can be found by either physical or statistical analysis, and it is frequently a combination in which a statistical relation is derived from values based on the knowledge of a physical process. Statistical methods such as regression are often used to establish such relations. The deterministic approach is stationary in time and space and must therefore be regarded as a global method reflecting the properties of the entire sample. The predictors may be other observed elements or other geographic parameters, such as elevation, slope or distance from the sea.

### 5.9.3 Spatial estimation methods

Spatial interpolation is a procedure for estimating the value of properties at unsampled sites within an area covered by existing observations. The rationale behind interpolation is that observation sites that are close together in space are more likely to have similar values than sites that are far apart (spatial coherency). All spatial interpolation methods are based on theoretical considerations, assumptions and conditions that must be fulfilled in order for a method to be used properly. Therefore, when selecting a spatial interpolation algorithm, the purpose of the interpolation, the characteristics of the phenomenon to be interpolated, and the constraints of the method have to be considered.

Stochastic methods for spatial interpolation are often referred to as geostatistical methods. A feature shared by these methods is that they use a spatial relationship function to describe the correlation among values at different sites as a function of distance. The interpolation itself is closely related to regression. These methods demand that certain statistical assumptions be fulfilled, such as that the process follows a normal distribution, it is stationary in space, or it is constant in all directions.

Even though it is not significantly better than other techniques, kriging is a spatial interpolation approach that has been used often for interpolating elements such as air and soil temperature, precipitation, air pollutants, solar radiation and winds. The basis of the technique is the rate at which the variance between points changes over space; this is expressed in a variogram. A variogram shows how the average difference between values at points changes with distance and direction between points. When developing a variogram, it is necessary to make some assumptions about the nature of the observed variation on the surface. Some of these assumptions concern the constancy of means over the entire surface, the existence of underlying trends, and the randomness and independence of variations. The goal is to relate all variations to distance. Relationships between a variogram and physical processes may be accommodated by choosing an appropriate variogram model (for example, spherical, exponential, Gaussian or linear).

Some of the problems with kriging are the computational intensity for large datasets, the complexity of estimating a variogram, and the critical assumptions that must be made about the statistical nature of the variation. This last problem is most important. Although many variants of kriging allow flexibility, the method was developed initially for applications in which distances between observation sites were small. In the case of climatological data, the distances between sites are usually large, and the assumption of smoothly varying fields between sites is often not realistic. In these cases, alternative approaches such as Climatologically Aided Interpolation (CAI) can also be explored.

Since meteorological and climatological fields such as precipitation are strongly influenced by topography, some methods, such as Analysis Using Relief for Hydrometeorology (AURELHY) and Parameter-elevation Regressions on Independent Slopes Model (PRISM), incorporate the topography into an interpolation of climatic data by combining principal components analysis, linear multiple regression and kriging. Depending on the method used, topography is described by the elevation, slope and slope direction, generally averaged over an area. The topographic characteristics are generally at a finer spatial resolution than the climate data.

Among the most advanced physically based methods are those that incorporate a description of the dynamics of the climate system. Similar models are routinely used in weather forecasting and climate modelling (see 6.3). As the computer power and storage capacity they require becomes more readily available, these models are being used more widely in climate monitoring, especially to estimate the value of climate elements in areas remote from actual observations (see 6.2.4 on reanalysis).

### 5.9.4 **Time series estimation**

Time series often have missing data that need to be estimated or values that must be estimated at time-scales that are finer than those provided by the observations. One or just a few observations can be estimated better than a long period of continuous missing observations. As a general rule, the longer the period to be estimated, the less confidence one can place in the estimates.

For single-station analysis, one or two consecutive missing values are generally estimated by simple linear, polynomial or spline approximations that are fitted from the observations just before and after the period to be estimated. The assumption is that conditions within the period to be estimated are similar to those just before and after the period to be estimated; care must be taken that this assumption is valid. An example of a violation of this assumption in the estimation of hourly temperatures is the passage of a strong cold front during the period to be estimated. Estimation of values for longer periods is usually accomplished with time series analysis techniques (see 5.5) performed on parts of the series without data gaps. The model for the values that do exist is then applied to the gaps. As with spatial interpolation, temporal interpolation should be validated to ensure that the estimated values are reasonable. Metadata or other corollary information about the time series is useful for determining the reasonableness.

At the regional to local scale, a range of tools, including stochastic weather generators, which use the statistical characteristics of observed meteorological time series to simulate weather time series, have been developed to infill and estimate meteorological variables in data-sparse regions.

### 5.9.5 **Validation**

Any estimation is based on some underlying structure or physical reasoning. It is therefore very important to verify that the assumptions made in applying the estimation model are fulfilled. If they are not fulfilled, the estimated values may contain errors that could be serious and lead to incorrect conclusions. In climatological analysis, model assumptions are often not met. For example, in spatial analysis, interpolating between widely spaced stations implies that the climatological patterns between stations are known and can be modelled. In reality, many factors (such as topography, local peculiarities or the existence of water bodies) influence the climate of a region. Unless these factors are adequately incorporated into a spatial model, the interpolated values will likely be wrong. In temporal analysis, interpolating over a large data gap implies that the values representing conditions before and after the gap can be used to estimate the values within the gap. In reality, the more variable the weather patterns are at a location, the less likely it is that this assumption will hold; consequently, the interpolated values could wrong.

The seriousness of any error of interpolation is related to the use of the data. Conclusions and judgments based on requirements for microscale (detailed information about a local area) will be much more affected by errors than those that are based on macroscale (general information for a large area). When estimating data, the sensitivity of the results to the use of the data should be considered carefully.

Validation is essential whenever estimation is performed. Split validation is a simple and effective technique. A large part of a dataset is used to develop the estimation procedures and a single, smaller subset of the dataset is reserved for testing the methodology. The data in the smaller subset are estimated with the procedures developed from the larger portion, and the estimated values are compared with the observed values. Cross-validation is another simple and effective tool for comparing various assumptions either about the models (such as the type of variogram and its parameters, or the size of a kriging neighbourhood) or about the data, using only the information available in the given sample dataset. Cross-validation is carried out by removing one observation from the data sample, and then estimating the removed value based on the remaining observations. This is repeated with the removal of a different observation from the sample, and repeated again, removing each observation in turn. The residuals between the observed and estimated values can then be further analysed statistically or can be plotted for visual inspection. Cross-validation offers quantitative insights into how any estimation method performs. An analysis of the spatial and temporal arrangement of the residuals often suggests further improvements of the estimation model.

## 5.10 EXTREME VALUE ANALYSIS

Many practical problems in climatology require knowledge of the behaviour of extreme values of some climatological elements. This is particularly true for the engineering design of structures that are sensitive to high or low values of meteorological or climatological phenomena. For example, high precipitation amounts and resulting stream flows affect sewerage systems, dams, reservoirs and bridges. High wind speed increases the load on buildings, bridges, cranes, trees and electrical power lines. Large snowfalls require that roofs be built to withstand the added weight. Public authorities and insurers may want to define thresholds beyond which damages resulting from extreme conditions become eligible for economic relief.

Design criteria are often expressed in terms of a return period, which is the mean interval of time between two occurrences of values equal to or greater than a given value. The return period concept is used to avoid adopting high safety coefficients that are very costly, but also to prevent major damage to equipment and structures from extreme events that are likely to occur during the useful life of the equipment or structures. As such equipment can last for years or even centuries, accurate estimation of return periods can be a critical factor in their design. Design criteria may also be described by the number of expected occurrences of events exceeding a fixed threshold.

### 5.10.1 Return period approach

Classical approaches to extreme value analysis represent the behaviour of the sample of extremes by a probability distribution that fits the observed distribution sufficiently well. The extreme value distributions imply assumptions such as stationarity and independence of data values. The three common extreme value distributions are Gumbel, Frechet and Weibull. The generalized extreme value (GEV) distribution combines these three under a single formulation, which is characterized by a model shape parameter.

The data that are modelled by an extreme value distribution are the maxima (or minima) of values observed in a specified time interval. For example, if daily temperatures are observed over a period of many years, the set of annual maxima could be represented by an extreme value distribution. Constructing and adequately representing a set of maxima or minima from subintervals of the whole dataset requires that the dataset be large, which may be a strong limitation if the data sample covers a limited period. An alternative is to select values beyond a given threshold, also known as Partial Duration Series. The generalized Pareto frequency distribution is usually suitable for fitting data beyond a threshold. Fitting techniques such as maximum likelihood and L-moments can also be used.

Once a distribution is fitted to an extreme value dataset, return periods are computed. A return period is the mean frequency with which a value is expected to be equalled or exceeded (such as once in 20 years). Although lengthy return periods for the occurrence of a value can be mathematically calculated, the confidence that can be placed in the results may be minimal. As a general rule, confidence in a return period decreases rapidly when the period is more than about twice the length of the original dataset.

Extreme climate events can have significant impacts on both natural and man-made systems, and it is therefore important to know if and how climate extremes are changing. Some types of infrastructure currently have little margin to buffer the impacts of climate change. For example, there are many communities in low-lying coastal zones throughout the world that are at risk from rising sea levels. Adaptation strategies to non-stationary climate extremes should take into account the decadal-scale changes in climate observed in the recent past, as well as future changes projected by climate models. Newer statistical models, such as the non-stationary generalized extreme value, have been developed to try to overcome some of the limitations of the more conventional distributions. As models continue to evolve and as their properties become better understood, they will likely replace the more common approaches to analysing extremes. The *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WMO/TD-No. 1500) provides more insight into how one should take account of a changing climate when assessing and estimating extremes.

### 5.10.2 **Probable maximum precipitation**

The probable maximum precipitation is defined as the theoretically greatest depth of precipitation for a given duration that is physically possible over a storm area of a given size under particular geographical conditions at a specified time of the year. It is widely used in the design of dams and other large hydraulic systems, for which a very rare event could have disastrous consequences.

The estimation of probable maximum precipitation is generally based on heuristic approaches, including the following steps:

(a) Use of a conceptual storm model to represent precipitation processes in terms of physical elements such as surface dewpoint, depth of storm cell, inflow and outflow;

(b) Calibration of the model using observations of storm depth and accompanying atmospheric moisture;

(c) Use of the calibrated model to estimate what would have occurred with maximum observed atmospheric moisture;

(d) Translation of the observed storm characteristics from gauged locations to the location where the estimate is required, adjusting for effects of topography, continentality, and similar non-meteorological or non-climatological conditions.

## 5.11 **ROBUST STATISTICS**

Robust statistics produce estimators that are not unduly affected by small departures from model assumptions. Statistical inferences are based on observations as well as on the assumptions of the underlying models (such as randomness, independence and model fit). Climatological data often violate many of these assumptions because of the temporal and spatial dependence of observations, data inhomogeneities, data errors and other factors.

The effect of assumptions on the results of analyses should be determined quantitatively if possible, and at least qualitatively, in an assessment of the validity of conclusions. The purpose of an analysis is also important. General conclusions based on large temporal or spatial scale processes with a lot of averaging and on a large dataset are often less sensitive to deviations from assumptions than more specific conclusions. Robust statistical approaches are often used for regression.

If results are sensitive to violations of assumptions, the analyst should flag this fact when disseminating the results to users. It may be also possible to analyse the data using other methods that are not as sensitive to deviations from assumptions, or that do not make any assumptions about the factors causing the sensitivity problems. Since parametric methods assume more conditions than non-parametric methods, it may be possible to re-analyse the data with non-parametric techniques. For example, using the median and interquartile range, instead of the mean and standard deviation, decreases sensitivity to outliers or to gross errors in the observational data.

## 5.12 STATISTICAL PACKAGES

Since most climatological processing and analyses are based on universal statistical methods, universal statistical packages are convenient computer software instruments for the climatologists. Several software products for universal statistical analysis are available on a variety of computer platforms.

Statistical packages offer numerous data management, analytical and reporting tools. A chosen package should have all the capabilities required to manage, process and analyse data, but not be burdened with unnecessary tools that lead to inefficiencies. Some of the basic tools are often included in a Climate Data Management System (see Chapter 3).

Basic data management tools provide a wide variety of operations with which to prepare the data for processing and analysis. These operations include sorting, adding data, sub-setting data, transposing matrices, arithmetic calculations and merging data. Basic statistical processing tools include the calculation of sample descriptive statistics, correlations, frequency tables and hypothesis testing. Analytical tools usually cover many of the needs of climate analysis, such as analysis of variance, regression analysis, discriminant analysis, cluster analysis, multidimensional analysis and time series analysis. Calculated results of analyses are usually put into resultant datasets and can be saved, exported and transformed, and thus used for any further analysis and processing.

Statistical packages contain graphical tools for creating two- and three-dimensional graphs, editing the graphs and saving them in the specific formats of the statistical packages or in standard graphical formats. Most packages can create scatter plots (two- and three-dimensional); bubble plots; line, step and interpolated (smoothed) plots; vertical, horizontal and pie charts; box and whisker plots; and three-dimensional surface plots, including contouring of the surface. Some packages contain tools for displaying values of some element on a map, but they should not be considered a replacement for a Geographical Information System (GIS). A Geographical Information System integrates hardware, software and data for capturing, managing, analysing and displaying all forms of geographically referenced information. Some GIS programs include geographical interpolation capabilities such as cokriging and geographically weighted regression tools.

Interactive analysis tools combine the power of statistical analysis and the ability to visually manage the conditions for any particular statistical analysis. Tools allow the visual selection of values to be included in or excluded from analyses, and recalculation based upon this selection. Such flexibility is useful in trend calculations when climate data series contain outliers and other suspicious points. These points can be interactively excluded from analysis based on a graph of the series, and trend statistics can be recalculated automatically. Options are usually available for analysing and displaying subgroups of data.

## 5.13 DATA MINING

Data mining is an analytic process designed to explore large amounts of data (big data) in search of consistent patterns or systematic relationships among elements, and then to validate the findings by applying the detected patterns to new subsets of data. It is often considered a blend of statistics, artificial intelligence and database research. It is rapidly developing into a major field, and important theoretical and practical advances are being made. Data mining is fully applicable to climatological problems when the volume of data available is large, and ways to search the significant relationships among climate elements may not be evident, especially at the early stages of analysis.

Data mining is similar to exploratory data analysis, which is also oriented towards the search for relationships among elements in situations when possible relationships are not clear. Data mining is not concerned with identifying the specific relations among the elements involved. Instead, the focus is on producing a solution that can generate useful predictions. Data mining takes a “black box” approach to data exploration or knowledge discovery and uses not only the traditional exploratory data analysis techniques, but also such techniques as neural networks, which can generate valid predictions but are not capable of identifying the specific nature of the interrelations among the elements on which the predictions are based.

## REFERENCES


World Meteorological Organization, 1966: *Some Methods in Climatological Analysis* (WMO/TN-No. 81, WMO-No. 199). Geneva.

———, 1990: *On the Statistical Analysis of Series of Observations* (WMO/TN-No. 143, WMO-No. 415). Geneva.

———, 1994: *Guide to the Applications of Marine Climatology* (WMO-No. 781). Geneva.

———, 2003: *Guidelines on Climate Metadata and Homogenization* (WCDMP-No. 53, WMO/TD-No. 1186). Geneva.

———, 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WCDMP-No. 72, WMO/TD-No. 1500). Geneva.


## FURTHER READING


Acharya, N., A. Frei, J. Chen, L. De Cristofaro and E.M. Owens, 2017: Evaluating Stochastic Precipitation Generators for Climate Change Impact Studies of New York City's Primary Water Supply. *Journal of Hydrometeorology*, 18 (3):879-896, doi:10.1175/JHM-D-16-0169.1.

Ailliot, P., D. Allard, V. Monbet and P. Naveau, 2015: Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique*, 156 (1):101-113.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *International Journal of Climatology*, 6:661–675.

Angel, J.R., W.R. Easterling and S.W. Kirtsch, 1993: Towards defining appropriate averaging periods for climate normals. *Climatological Bulletin*, 27:29–44.

Beljaars, A.C.M., L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N.A. Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen, 2005: The ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 131:2961–3012.

Bénichou, P. and O. Le Breton, 1987: Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie*, 7(19):23–34.

Burrough, P.A., 1986: *Principles of Geographical Information Systems for Land Resources Assessment*. Monographs on Soil and Resources Survey No. 12. Oxford, Oxford University Press.

Cook, E.R., K.R. Briffa and P.D. Jones, 1994: Spatial regression methods in dendroclimatology: A review and comparison of two techniques. *International Journal of Climatology*, 14:379–402.

Daley, R., 1993: *Atmospheric Data Analysis*. New York, Cambridge University Press.

Dalgaard, P., 2002*: Introductory Statistics with R*. New York, Springer.

Daly, C., R.P. Neilson and D.L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33:140–158.

Della-Marta, P.M. and H. Wanner, 2006: Method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate,* 19:4179–4197.

Deutsch, C.V. and A.G. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide, s*econd edition. Oxford, Oxford University Press.

Dixon, K.W. and M.D. Shulman, 1984: A statistical evaluation of the predictive abilities of climatic averages. *Journal of Climate and Applied Meteorology*, 23:1542–1552.

Domonkos, P. and J. Coll, 2017: Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. *International Journal of Climatology*, 37(4):1910-1921.

Donat, M.G., L.V. Alexander, H. Yang, I. Durre, R. Vose, R.J.H. Dunn, K.M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A.M.G. Klein Tank, A.C. Kruger, J. Marengo, T.C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A.S. Elrayah, S.S. Sekele, A.K. Srivastava, B. Trewin, C. Villarroel, L.A. Vincent, P. Zhai, X. Zhang, S. Kitching, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres,* 118:2098–2118, doi:10.1002/jgrd.50150.

Environmental Systems Research Institute (ESRI), 2008: *ArcGIS Geographical Information System*. Redlands.

Fisher, R.A. and L.H.C. Tippet, 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190.

Frich, P., L.V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A.M.G. Klein Tank and T. Peterson, 2002: Global changes in climatic extreme events during the second half of the 20th century. *Climate Research*, 19:193–21.

Gabor, D, 1946: Theory of communication. *J. IEEE*, 93:429–457.

Gandin, L., 1963: Objective Analysis of Meteorological Fields. Leningrad, Hydrometeoizdat, in Russian (English translation: Israel Program for Scientific Translation, Jerusalem, 1965).

Haylock, M.R., N. Hofstra, A.M.G. Klein Tank, E.J. Klok, P.D. Jones and M. New, 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113: D20119, doi:10.1029/2008JD010201.

Heino, R., 1994: *Climate in Finland During the Period of Meteorological Observations*. Contributions, No. 12. Helsinki, Finnish Meteorological Institute.

Intergovernmental Panel on Climate Change (IPCC), 1997: *An Introduction to Simple Climate Models used in the IPCC Second Assessment Report: IPCC Technical Paper II* (J.T. Houghton, L.G. Meira Filho, D.J. Griggs and K. Maskell, eds.). Geneva.

International ad hoc Detection and Attribution Group (IDAG), 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate*, 18:1291–1314.

Jenkinson, A.F., 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171.

Jones, P.D., S.C.B. Raper, B. Santer, B.S.G. Cherry, C. Goodess, P.M. Kelly, T.M.L. Wigley, R.S. Bradeley and H.F. Diaz, 1985: *A Grid Point Surface Air Temperature Data Set for the Northern Hemisphere*, DOE/EV/10098-2. Washington, DC, United States Department of Energy.

Jones, P.D., E.B. Horton, C.K. Folland, M. Hulme and D.E. Parker, 1999: The use of indices to identify changes in climatic extremes. *Climatic Change*, 42:131–149.

Jureckova, J. and J. Picek, 2006: *Robust Statistical Methods with R*. Boca Raton, Chapman and Hall/CRC.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deavan, M. Iredell, S. Saha, G. White, J. Woolen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne and D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77:437–471.

Kaplan, A., Y. Kushnir, M. Cane and M. Blumenthal., 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research*, 102:27835–27860.

Karl, T.R., C.N. Williams Jr and P.J. Young, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. *Journal of Climate and Applied Meteorology*, 25:145–160.

Kharin, V.V. and F.W. Zwiers, 2005: Estimating extremes in transient climate change simulations. *Journal of Climate*, 18:1156–1173.

Kohler, M.A., 1949: Double-mass analysis for testing the consistency of records for making adjustments. *Bulletin of the American Meteorological Society*, 30:188–189.

Mann, M. E., 2004: On smoothing potentially non-stationary climate time series. *Geophysical Research Letters,* 31:L07214, doi:10.1029/2004GL019569.

Mehan, S., T. Guo, M.W. Gitau, D.C. Flanagan, 2017: Comparative study of different stochastic weather generators for long-term climate data simulation. *Climate*, 5 (2), article No. 26, doi:10.3390/cli5020026.

National Climatic Data Center (NCDC), 2006: *Second Workshop on Climate Data Homogenization*, National Climatic Data Center (Asheville, North Carolina, 28–30 March 2006). Jointly organized by the Climate Research Division, Environment Canada, and the National Climatic Data Center, NOAA.

Palus, M. and D. Novotna, 2006: Quasi-biennial oscillations extracted from the monthly NAO index and temperature records are phase-synchronized. *Nonlinear Processes in Geophysics*, 13:287–296.

Peterson, T.C., D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E.J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker, 1998: Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18:1493-1517.

Peterson, T.C., C. Folland, G. Gruza, W. Hogg, A. Mokssit and N. Plummer, 2001: *Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998–2001*. International CLIVAR Project Office (ICPO) Publication Series, 48. Southampton, United Kingdom.

Peterson, T. C. et al., 2002: Recent changes in climate extremes in the Caribbean Region. *Journal of Geophysical Research*, 107(D21):4601.

Philipp, A., P.M. Della-Marta, J. Jacobeit, D.R. Fereday, P.D. Jones, A. Moberg and H. Wanner, 2007: Long-term variability of daily North Atlantic–European pressure patterns since 1850 classified by simulated annealing clustering. *Journal of Climate*, 20:4065–4095.

Ribeiro, S., J. Caineta and A. Costa, 2016: Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth*, 94: 167-179.

Ribeiro, S., J. Caineta, A.C. Costa, R. Henriques, 2017: GSIMCLI: A geostatistical procedure for the homogenisation of climatic time series. *International Journal of Climatology*, 37 (8):3452-3467.

Schnase, J.L., T.J. Lee, C.A. Mattmann, C.S. Lynnes, L. Cinquini, P.M. Ramirez, A.F. Hart, D.N. Williams, D. Waliser, P. Rinsland, W.P. Webster, D.Q. Duffy, M.A. Mcinerney, G.S. Tamkin, G.L. Potter and L. Carrier, 2016: Big data challenges in climate science: Improving the next-generation cyberinfrastructure. *IEEE Geoscience and Remote Sensing Magazine*, 4 (3):10-22, doi:10.1109/MGRS.2015.2514192.

Sen, P.K., 1968: Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63:1379–1389.

Sensoy, S., T.C. Peterson, L.V. Alexander and X. Zhang, 2007: Enhancing Middle East climate change monitoring and indices. *Bulletin of the American Meteorological Society*, 88:1249–1254.

Shabbar, A. and A.G. Barnston, 1996: Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Monthly Weather Review,* 124(10):2370–2385.

Torrence, C. and G.P. Compo, 1998: A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78.

Ulbrich, U. and M. Christoph, 1999: A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing. *Climate Dynamics*, 15:551–559.

Uppala, S.M., P.W. Kållberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P, Allan, E. Andersson, K. Arpe, M.A. Balmaseda, Vincent, L.A., X. Zhang, B.R. Bonsal and W.D. Hogg, 2002: Homogenization of daily temperatures over Canada. *Journal of Climate*, 15:1322–1334.

von Storch, H. and F.W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge, Cambridge University Press.

Wackernagel, H., 1998: Kriging, cokriging and external drift. In: *Workshop on Dealing with Spatialisation* (B. Gozzini and M. Hims, eds.). Toulouse, European Union COST Action 79.

Wahba, G. and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Review*, 108:36–57.

Wang, X.L. and V. R. Swail, 2001: Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *Journal of Climate*, 14:2204–2221.

Wang, X.L., 2008: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *Journal of Applied Meteorology and Climatol*ogy, 47:2423–2444.

Wilks, D.S., 2002: *Statistical Methods in the Atmospheric Sciences*. Second edition. New York, Academic Press.

Willmott, C.J. and S. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air-temperature. *International Journal of Climatology*, 15 (2):221-229.

Yozgatligil, C. and C. Yazici, 2016: Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology*, 36 (1):62-81, doi:10.1002/joc.4329.

Yue, S., P. Pilon and G. Cavadias, 2002: Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259:254–271.

Zhang, X., E. Aguilar, S. Sensoy, H. Melkonyan, U. Tagiyeva, N. Ahmed, N. Kutaladze, F. Rahimzadeh, A. Taghipour, T.H. Hantosh, P. Albert, M. Semawi, M. Karam Ali, M. Halal Said Al-Shabibi, Z. Al-Oulan, Taha Zatari, I. Al Dean Khelet, S. Hammoud, M. Demircan, M. Eken, M. Adiguzel, L. Alexander, T. Peterson and Trevor Wallis, 2005: Trends in Middle East climate extremes indices during 1930–2003. *Journal of Geophysical Research*, 110:D22104, doi:10.1029/2005JD006181.

# CHAPTER 6. CLIMATE PRODUCTS AND THEIR DISSEMINATION

## 6.1 GENERAL GUIDELINES

Climate products are information packages that include data, summaries, tables, graphs, maps, reports and analyses. Spatial distributions may be shown on maps. More complex products, such as climate atlases or analyses, may combine several kinds of visualization with descriptive text. There may also be databases with software tools that allow online customers to produce statistics and visualizations according to their own needs.

Products and the data on which they are based should be of the highest quality possible within the time constraints for providing the information. There has been a strong and increasing requirement for climate-related products to be provided as quickly as possible after the aggregating period. Maintaining the quality standards for such products is a concern. The short time between observation and delivery to a user leaves little or no time for quality control of the data other than that which can be done automatically. At the very least some basic checks should be made as the data are received (see Chapter 3). Users must be alerted to possible problems concerning the data, and since these products are usually automatically delivered, such alerts should be included with the products. A proper quality assurance system will provide a framework that allows this information to be handled along with the data.

Products concerning historical data should be of higher quality than those using very recent data. All data that contribute to the climate record should be checked for random and systematic errors, homogeneity, spatial representativeness and gaps in time series. For products such as climatic atlases or technical regulations, data should be for a standard reference period (see 4.8). Frequent revisions based on new periods of record should be avoided. If some content of a product is not stable over a long period of time, there needs to be additional information describing the nature of the variability or change.

The value of historical and statistical climatological data tables can usually be improved by the inclusion of a supporting text that helps the user to interpret the data and emphasizes the more important climatological elements. In all publications, sufficient information and data must be included regarding the location and elevation of the observing stations, the homogeneity of the data from all stations, the periods of record used and the statistical or analytical procedures employed.

The display of products should be checked carefully before they are made accessible to potential users. For example, climatic maps should be well designed with carefully chosen colours and scales, clear titles and notations on the map of what is being analysed, identification of the data period of record, and a listing of the responsible organizations. There should be reasonable consistency among maps (in terms of colours, layout and data) to allow for easy comparisons.

Consultation with everyone who is affected by environmental information services is encouraged. Input from interested parties should be considered when creating, modifying or discontinuing products and services.

### 6.1.1 Climatological data periodicals

A periodical climatological publication is one that is scheduled for preparation and publication on a routine basis over set time intervals. Most climatological data periodicals are issued on either a monthly or annual basis. Some services, however, also publish periodicals at different intervals such as a week or a season. Weekly or monthly publications are issued immediately after the close of the period in question and usually contain recent data that have not undergone complete quality control procedures. These periodicals contain timely data that can be of great importance to various economic, social and environmental sectors, and so publication is valuable even though the data may contain a few errors and omissions. Quarterly or seasonal data periodicals are often issued to disseminate summarized seasonal data such as winter snowfall, growing-season precipitation, summer cooling degree-days and winter degree-days.

Most National Meteorological and Hydrological Services (NMHSs) issue monthly bulletins containing data from a selection of stations within particular areas or states or the country as a whole. When issued a week or two after the end of each month, these periodicals will usually contain recent data that might not have undergone full quality control, but if issued a month or more afterwards, all data should meet the normal quality control standards for historical climatological data. Maximum and minimum temperature and total precipitation for each day should be listed, as well as perhaps temperatures at fixed hours, together with the associated humidity values. Daily mean wind speed and prevailing direction, duration of bright sunshine, or other locally important data (such as heating, cooling and growing degree-days) could also be included. Monthly averages, extremes and other statistical data from all stations should also be included when available.

While most monthly bulletins contain only surface climatological data, some NMHSs include a selection of basic data from upper-air stations or issue separate monthly bulletins containing upper-air data. In such monthly bulletins, daily and monthly mean data are usually published for the standard pressure surfaces. The data usually include altitude (in geopotential meters), temperature, humidity, and wind speed and direction for one or two scheduled ascents each day.

Some of the most useful climatological publications are those containing simple tables of monthly and annual values of mean daily temperature and total precipitation. Such tables are prepared by NMHSs and are made available either in manuscript or electronic format. The services should publish, at least once a decade, a comprehensive set of statistical climatological data for a selection of representative stations.

Periodicals sponsored by WMO include data from Members. Examples are *Monthly Climatic Data for the World* (data from all CLIMAT stations), *World Weather Records* (single-station, historical, monthly and annual values of station pressure, sea-level pressure, temperature and precipitation), and *Marine Climatological Summaries* (monthly, annual and decadal climatological statistics and charts for the oceans).

Since 1993, WMO, through the Commission for Climatology and in cooperation with its Members, has also issued annual statements on the status of the global climate to provide credible scientific information on climate and its variability.

A useful monthly publication, especially in relation to El Niño-Southern Oscillation (ENSO) diagnostics and other climate teleconnections, is the *Climate Diagnostics Bulletin* published by the US National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center.

### 6.1.2 Occasional publications

Unlike climate data periodicals, which are produced to a schedule, occasional publications are produced as the need arises. They are in a form that will satisfy a large number of users for a considerable time, so they will not need frequent updating. Occasional publications are designed for those users who need information when planning for capital investments or designing equipment and buildings to last for decades and centuries; for members of the general public whose interests are academic or casual; and for researchers in the atmospheric and oceanic sciences. They are also designed to summarize or explain unusual events, such as extreme weather, and to describe or update an important predicted event such as a strong El Niño. The content and format of a specific occasional publication must reflect the interests and needs of the users for whom it is published.

Long-term, continuous and homogeneous series of data are of great value for comparative climatological studies and for research on climatic fluctuations, trends and changes. Several NMHSs have published such series for a selection of stations where observational practices and the environmental surroundings have remained essentially unchanged over long periods of time.

Data series most commonly available and needed are those of temperature and precipitation, although data for wind, pressure, bright sunshine, cloudiness and other climatic elements might also be published. Some NMHSs include historical climatological data series in yearbooks or other annual bulletins. Monographs on the climate of a country or area are valuable to a wide range of users and should be published and updated periodically. It is recommended that the publications and data be available also in electronic format for ease of access and exchange.

The collection of maps in atlas format is another valuable occasional publication. Legends and captions on climatic maps should include precise information regarding the element mapped, some indication of the number of stations from which data have been obtained, and the period of record used to generate each map or diagram.

### 6.1.3 Standard products

Although products specifically tailored to individual users may be the best for those users, it is usually beneficial to develop a standard product that can be used by a wide range of users. For example, both energy management entities and fruit growers can make use of a degree-day product. When the content, format and design of a product are carefully chosen, the development costs can be spread across many users. Such standard products fill the gap between the climate data periodicals and those tailored to individual users. Standard products should be locally developed to meet the needs of groups of users.

Increasingly, products are being requested and delivered using the Internet. The user interface for these systems can be considered another product of the climate service, and standardizing that interface can be seen as enhancing the quality and utility of the product.

### 6.1.4 Specialized products

It is often necessary to develop products that are specific to an individual user or sector. The particular requirements of one user group do not always match the requirements of other groups, so the expense of publishing the product for general availability is not warranted.

Such applied climatological products are tailored to the needs of a particular user or user group. These products provide a bridge between the observed data and the specific requirements of a user; they transform the observations into a value-added product for particular recipients. Developing these products involves analysing the data and presenting the information with a focus on the specifications that will enable the user to gain optimum benefit from the application of the information. The use of the product usually dictates the types of analysis and data transformation that need to be performed and the methods used to deliver the product.

The climate service should be able to accept requests for specialized products and develop the products to the satisfaction of the users, which will require all of the skills of user interaction and marketing already discussed. Although the product may not be published for a general audience, the users will expect at least the same level of quality, both in content and presentation.

An example of an application-driven product can be found in the requirement by a fruit grower for daily degree-hour data for pesticide management of fire blight disease. When only daily maximum and minimum temperatures are available for the locations of interest, degree-days can be calculated from the average of the maximum and minimum values. Since degree-hours are required but not available, an analysis is necessary to develop relationships between degree-days calculated from daily temperature extreme data and degree-hours. The conditions for which the relationships are valid, and the degree of error in the relationships, must also be assessed. Once the relationships are established, the user can be given a product that contains degree-hours, even though degree-hours are not measured directly.

Flood analysis is another example. Flooding is a natural feature and varies in scale from water running off a saturated hillside to large rivers bursting their banks. The impacts of floods range from waterlogged fields and blocked roads to widespread inundation of houses and commercial property and, occasionally, loss of life. Flood frequency estimates are required for the planning and assessment of flood defences; the design of structures such as bridges, culverts and reservoir spillways; and the preparation of flood risk maps for the planning of new developments and for insurance interests. A product that provides the probability of observed precipitation amounts is a necessary component in the development of flood frequency estimates. Developing the precipitation risk information involves the value-added statistical analysis (see Chapter 5) of the observed precipitation data that are usually presented in standard summaries. If the resulting risk analyses will be of use to a number of different users, a general publication may be warranted.

### 6.1.5 Climate monitoring products

Monitoring climate variability around the world is a goal of the World Climate Data Monitoring Programme (WCDMP). Maintenance and accessibility of climate data and information by a climate service supports this WCDMP objective. For monitoring and diagnosing the climate of a country, it is necessary to understand current climate conditions in the country as part of the global climate system. In addition to monitoring local climates for national interests and relating current episodes to historical patterns, the climate service should aim to place the local variations within a larger regional and even global context.

Observation and monitoring of climate may be conducted by more than one agency in a country. When an agency publishes observational data, analytical results and statistical data, the products are usually presented in formats suited to the agency's own purposes. Hence, the products may not necessarily be appropriate for use by other agencies. In addition, it may not be easy for individual users to choose products for their own needs from among the climate monitoring products distributed by various agencies, or to consolidate a disparate set of products in order to grasp an understanding of the whole climate system. For many users it is also difficult to understand the connection between features of the global climate system and the current climate conditions within their own country. Thus, the climate service should process its own data and analysis results and, where possible, compile them together with the material of other agencies into a set of products that can be promptly disseminated with each agency's views on current climate conditions.

If the necessary data are not available within a given country, the relevant NMHS should obtain regional or global data and analyses from foreign or international agencies and process the information into a form suitable for local to national use. The NMHS should, however, add its own views to these global analyses about the connection between the local climate conditions and the large-scale climatic fields. Monitoring activities require that the climate service develop expertise in analysing the state of both past and current climate and global to regional teleconnections, and provide summarized information to both public and private sector users. Good monitoring products are essential for climate predictions and updates.

### 6.1.6 Indices

Presentation of historical climate patterns to the user in a simple and readily understandable form may often be accomplished with indices (see 4.4.6). Climate indices are widely used to characterize features of the climate for climate prediction and to detect climate change. They may apply to individual climatological stations or describe some aspect of the climate of an area. Indices usually combine several elements into characteristics of, for example, droughts, continentality, phenological plant phases, heating degree-days, large-scale circulation patterns and teleconnections. When providing information to users, it is often necessary for the climate service to interpret the meaning of an index value, changes in values over time, and sometimes calculation procedures. Examples of indices are the ENSO Index; the North Atlantic Oscillation Index; descriptors such as the moisture availability index, used for deriving crop planning strategies; agrometeorological indices such as the Palmer Drought Severity Index, aridity index and leaf area index, which are used for describing and monitoring moisture availability; and the mean monsoon index, which summarizes areas of droughts and floods. The construction and evaluation of indices specific to climate change detection, climate variability and climate extremes are ongoing processes, as discussed in *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WMO/TD-No. 1500, WCDMP-No. 72).

## 6.2 GRIDDED DATA

Gridded climate data products are values of surface or upper-air climate variables (for example, air temperature, atmospheric moisture or sea surface temperature) or indices (for example, number of frost days), arranged on a regular grid with coverage ranging from the local to regional to global. In addition to the scale of coverage, the resolution of gridded data can vary from as little as a few square metres in the case of sub-urban datasets to 200-300km as found in global scale datasets. Similarly, temporal resolution may vary from the sub-hourly to annual timescale.

Because gridded climate data is in essence an alternative to instrumental measurements, irrespective of whether the gridded data has been derived from original observations using interpolation techniques (see 5.9) or from the output of numerical or statistical climate models, it is often subject to validation. This is achieved through a comparison with observations from surface or upper-air climate stations that either coincide with a particular grid point or are close by.

A range of organisations are involved in the production of gridded climate datasets, for example national or regional climate centres and university research groups. Some datasets are updated regularly or periodically, while some are static or contain times series of climate variables.

Gridded datasets facilitate the spatial analysis of climate variables and the static or dynamic visualisation of climate patterns and trends.

### 6.2.1 Gridded datasets based on observations

An acknowledged frustration in climatological analysis is the patchiness of data over both space and time due to an uneven geographical and temporal distribution of climate observations. Over the last decade or so not only have reanalysis systems opened up new possibilities for climate analysis (see 6.2.4) but the availability of observation-based gridded climate datasets has helped address some of the challenges posed by the incompleteness of observed data. While some of the observed gridded climate datasets are based purely on surface observations others are blended datasets including observations from both surface and satellite-based platforms. Typically observation-based gridded climate datasets are comprised of values interpolated from stations for which data has been adjusted and homogenized.

Some examples of global and regional observation-based gridded datasets are provided in Table 6.1.

**Table 6.1. Examples of global and regional observation-based gridded datasets**

| *Global dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| Climate Prediction Centre (CPC) Merged Analysis of Precipitation (CMAP)[a] | Monthly and pentad global gridded precipitation means | Monthly values: January 1979–September 2017<br><br>Pentad values: January 1979–27 December 2016<br><br>Long-term monthly means derived from data for 1981–2010 | 2.5 degree latitude x 2.5 degree longitude global grid |

| *Global dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| CPC Soil Moisture[a] | Soil moisture | Monthly means: January 1948 through August 2017<br><br>Long-term monthly means derived from data for 1981–2010 | 0.5 degree latitude x 0.5 degree longitude global grid |
| CRU TS v. 4.01[b] | pre, tmp, tmx, tmn, dtr, vap, cld, wet, frs, pet | Time series 1901–2016 | All land areas, 0.5 degree latitude x 0.5 degree longitude |
| CRU CY v. 4.01 longitude global grid[b] | pre, tmp, tmx, tmn, dtr, vap, cld, wet, frs, pet | Mean values: 1901–2015 | Countries |
| GHCN[c] Gridded Land Precipitation (V2) and Temperature Anomalies (V3) | Monthly precipitation and temperature | Monthly anomalies: from 1900 to May 2015 (precipitation)<br><br>From 1880 to July 2016 (temperature)<br><br>Monthly means from 1900 to May 2015 (precipitation only) | 5.0 degree latitude x 5.0 degree longitude global grid |
| GHCN_CAMS[d] Gridded 2m Temperature (Land) | Analysed global land surface temperatures | Monthly means from 1948 to September 2017<br><br>Long-term monthly means using 1981–2010 data | 0.5 degree latitude x 0.5 degree longitude global grid |
| Global Precipitation Climatology Centre (GPCC)[e] | Monthly precipitation dataset from station | From 1901to present<br><br>Monthly values: January 1901 through 2013 (full V7)<br><br>Monthly values : January 2007 through near present (monitoring)<br><br>Monthly values: January 2014 through near present (first guess) | 0.5 degree latitude x 0.5 degree longitude global grid<br><br>1.0 degree latitude x 1.0 degree longitude global grid<br><br>2.5 degree latitude x 2.5 degree longitude global grid |
| NOAA Interpolated Outgoing Longwave Radiation (OLR)[a] | Gridded daily and monthly OLR data | Monthly values: June 1974–December 2013<br><br>Daily values: June 1974–31 December2013<br><br>Long-term means for monthly and daily values: 1981–2010 | 0.5 degree latitude x 2.5 degree longitude global grid |
| Kaplan Extended SST (Sea Surface Temperature) V2[a] | Gridded global SST anomalies from 1856 to present, derived from UK Met Office SST data | January 1856–September 2017 | 5.0 degree latitude x 5.0 degree longitude global grid |

| *Global dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| NOAA Extended Reconstructed SST V3b[a] | A global monthly SST analysis from 1854 to present, derived from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) with missing data filled in using statistical methods | Monthly values: January 1854–present<br><br>Long-term monthly means derived from data for 1971–2000<br><br>Long-term monthly means derived from data for 1981–2010 | 2.0 degree latitude x 2.0 degree longitude global grid |
| NOAA Extended Reconstructed SST V5[a] | A global monthly SST analysis from 1854 to present, derived from ICOADS with missing data filled in using statistical methods | Monthly values: January 1854–present<br><br>Long-term monthly means derived from data for 1981–2010 | 2.0 degree latitude x 2.0 degree longitude global grid |
| CPC Global Unified Gauge-based Analysis of Daily Precipitation[a] | Global unified daily gauge-based analysis of precipitation | Daily values: 1 January 1979 to present<br><br>Long-term means of daily and monthly values for 1981–2010 | 0.50 degree latitude x 0.50 degree longitude grid |
| CPC Global Daily Temperature[a] | Global daily land temperature | Daily values: 1 January 1979– present<br><br>Long-term means of daily and monthly values for 1981–2010 | 0.50 degree latitude x 0.50 degree longitude grid |
| NOAA Optimum Interpolation (OI) SST V2[a] | Sea surface temperature | Weekly means: 29 October 1981–28 December 1989<br><br>Monthly means: December 1981–September 2017<br><br>Long-term monthly means derived from data for 1961–1990<br><br>Long-term monthly means derived from data for 1971–2000 | 1.0 degree latitude x 1.0 degree longitude global grid |
| NOAA Outgoing Longwave Radiation–Daily Climate Data Record (OLR–Daily CDR): PSD Interpolated Version[a] | Gridded daily 1x1 OLR interpolated | Daily values: 1 January 1979–12 January 2012<br><br>Monthly values: 1 January 1979–12 January 2012<br><br>Long-term means for 1981–2010 | 1.0 degree latitude x 1.0 degree longitude global grid |

| *Regional dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| Canadian Gridded Temperature and Precipitation Anomalies (CANGRD)[f] | Gridded temperature and precipitation anomalies | Annual, seasonal and monthly temperature, and precipitation anomalies: 1901–2003 | 10 km resolution |

| *Regional dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| CPC Hourly US Precipitation[a] | Precipitation | Hourly values: January 1948–September 2002 | 2.0 degree latitude x 2.5 degree longitude US grid (33x21)<br><br>20N–60.0N, 220.0E–297.5E |
| E (Europe)-OBS gridded dataset V16[g] | Mean, minimum and maximum temperature, precipitation sum and averaged sea level | Daily values: 1 January 1950–31 August 2017 | 0.25 and 0.5 degree regular latitude-longitude grid, as well as on a 0.22 and 0.44 degree rotated pole grid |
| Africa Rainfall Climatology ARC2[h] | Rainfall | Daily values: since 1 January 1960, 12:00:00 ordered from 1 January 1983–1 November 2018 | From 20W to 55E by 0.1 degree |
| PSD South American Daily Gridded Precipitation [a] | Precipitation | Daily values: January 1940 through April 2012 | Land areas only 60°S–15°N 85°W–30°W 1.0 and 2.5 degree grids |
| Australia gridded rainfall variability[i] | Rainfall | Three-monthly and annual rainfall, variability indices | 0.25°/25 km grids |
| Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) (APHRO_MA/ME/RU_V 1101R1)[j] | Precipitation | Daily values: 1951–2007 | It varies depending on product |

[a] National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory, Physical Sciences Division (PSD)
[b] Climatic Research Unit (CRU), University of East Anglia
[c] Global Historical Climatology Network (GHCN), National Centers for Environmental Information, NOAA
[d] Climate Anomaly Monitoring System (CAMS)
[e] Deutscher Wetterdienst, Germany
[f] Environment and Climate Change Canada
[g] European Climate Assessment & Dataset (ECA&D)
[h] NOAA Climate Prediction Center
[i] Australian Bureau of Meteorology
[j] Research Institute for Humanity and Nature, and the Meteorological Research Institute of Japan Meteorological Agency

### 6.2.2 **Gridded datasets based on climate models**

There is a plethora of both numerical and statistical climate models of past, current and future climate at a variety of geographical and temporal scales, all of which have model output in the form of gridded data associated with them. Although model data are a numerical or statistical representation of actual climate conditions, one advantage of climate modelling is the vast range of climate variables, beyond those observed with traditional instruments, which are possible as output and therefore available as gridded datasets. Some examples of gridded datasets based on climate models are provided in Table 6.2.

**Table 6.2. Examples of global and regional gridded datasets based on climate models**

| *Dataset* | *Variables* | *Temporal coverage* | *Spatial coverage* |
|---|---|---|---|
| Coupled Model Intercomparison Project 5 (CMIP5) [a] | Comprehensive set of surface and upper-air variables from 62 climate models | Daily, monthly, long-term (century timescale) and near-term (10–30yr) climate simulations | Global, grid sizes ranging from 1.5 to 3.5 degrees |
| Coordinated Regional Downscaling Experiment (CORDEX)[b] | Comprehensive set of surface and upper-air variables from 54 regional climate models | Three-hourly, six-hourly, daily, monthly, long-term (century timescale) and near-term (10-30yr) climate simulations | 14 regions at 50 km/0.5 degrees grid size |

[a] Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, United States

[b] Swedish Meteorological and Hydrological Institute, Norrköping

### 6.2.3 **Hindcast data**

A hindcast is a prediction of a past event in which only observations prior to the event are fed into the system used to make the prediction (IPCC, 2013; WMO METEOTERM). Occasionally, hindcasts are referred to as "reforecasts", which are standardized datasets for use in improving weather forecasts (Hamill et al., 2006). They are retrospective weather forecasts generated with the same version of the numerical weather forecasting model used for the real-time weather forecasts being examined. A sequence of hindcasts can be used to calibrate the forecast system or provide a measure of the average capability that the forecast system has exhibited in the past as a guide to the performance that might be expected in the future.

The quality of a forecast (or prediction) measures the success of a prediction against observation-based information. Forecasts of past events, i.e. hindcasts or retrospective forecasts, may be analysed to give an indication of the quality that may be expected of future forecasts for a particular variable at a particular location. The importance of the hindcast or retrospective forecast as a tool to improve the knowledge of physical mechanisms driving specific processes or the identification of systematic biases in modelling experiments is illustrated by a few examples.

As an example of application of the technique, Katragkou et al. (2015) present six hindcast Weather Research and Forecasting (WRF) model simulations for the European Coordinated Regional Downscaling Experiment (EURO-CORDEX) domain with different configurations in microphysics, convection and radiation for the time period 1990–2008, with the aim of identifying systematic biases and areas of large uncertainties in present European climate, and relating them to specific physical processes (for example, cloud–radiation or land–atmosphere interactions).

Another study by Dake et al. (2004) presents retrospective forecasts of the inter-annual climate fluctuations in the tropical Pacific Ocean for the period 1857-2003, using a coupled ocean–atmosphere model. The model successfully predicts all prominent El Nino events within this period at lead times of up to two years. These results show that self-sustaining internal dynamics rather than stochastic forcing is controlling the evolution of the phenomenon.

As for application, hindcast studies of wave energy in specific coastal areas over shorter periods help to determine the variability and trends across areas, leading to the wave climate characterization which is an essential tool for wave power assessment.

### 6.2.4 **Reanalysis**

In operational numerical weather analysis and prediction, “analysis” refers to the process of creating an internally consistent representation of the environment on a four-dimensional grid. The time-critical nature of weather prediction means that the initializing analysis must usually begin before all observations are available. Reanalysis uses the same process (and often the same systems), but as it is done weeks or even years later, it is able to use a more complete set of observations. These reanalysis systems generally incorporate a prediction model that provides information on how the environment is changing with time, while maintaining internal consistency. Unlike the "analysis" in operational weather prediction, in which the models are constantly updated to incorporate the latest research advances, the "reanalysis" is performed with a fixed modelling system throughout the period of reanalysis to prevent the inhomogeneities that generally exist in the operational dataset because of model differences over time.

The output of a reanalysis is on a uniform grid and no data are missing. It is important to note that the reanalysis values are not "real" data, rather they are estimates of real data based on unevenly distributed observational data. The result is an integrated historical record of the state of the atmospheric environment for which all the data have been processed in the same manner. Reanalysis outputs are often used in place of observational data, but this must be done with care. Although the analysis algorithms will make good use of observations when available, in regions where observations are scarce the reanalysis grid will be strongly influenced by the prediction model. For reanalysis projects that span decades, there is generally a lot of heterogeneity in the type and coverage of data throughout the period, such as between the pre- and post-satellite periods. Further, the relative influence of the observations and the model is different for different climatic variables; certain variables are strongly influenced by the observational data used, while some are purely model-derived. These aspects should be carefully considered when interpreting the reanalysis data products. For example, reanalyses of dynamical variables are far better than reanalyses of precipitation, partially because processes leading to precipitation are not well represented in the models.

The limitations of reanalysis outputs are most obvious in areas with complex orography (in particular, in mountain regions), as well as in other areas when the assimilation and processing schemes are unable, because of smoothing, to reproduce real atmospheric processes with high spatial and temporal gradients. Also, there still remains the issue of localizing to spatial and temporal scales finer than the reanalysis grid. There are ongoing efforts to perform "regional reanalysis" using more local observational data with higher-resolution limited area models. As with any other analysis technique, validation of models, quality assurance and indicators of error are necessary to properly interpret the results.

The information from not only atmospheric sciences, but also oceanography, hydrology and remote-sensing, is used to create environmental databases from which systematic changes can be better assessed. Currently, the main global-scale reanalysis databases are those of the National Center for Atmospheric Research and National Centers for Environmental Prediction in the United States, the European Centre for Medium-Range Weather Forecasts, and the Japan Meteorological Agency (Table 6.3). All of these reanalysis efforts have found wide use in climate monitoring, climate variability studies and climate change prediction. It is important to assess the relative skill of the reanalysis techniques in representing the observed features in a given region before using their data for further climatological studies. Greater understanding of the physical, chemical and biological processes that control the environment, combined with data from a range of sources that go well beyond the atmospheric sciences, should promote improvements in reanalysis databases. As the numerical models become more complete, and as computer technology allows for higher resolution, more accurate and comprehensive reanalysis products will emerge.

**Table 6.3. Global and regional reanalysis systems**

| *Reanalysis system* | *Origin* | *Period* | *Website* |
|---|---|---|---|
| ASR[a] | The Ohio State University | 2000–2012 | http://rda.ucar.edu/datasets/ds631.4/ |
| CFSR[b] | NOAA/NCEP (CDAS-T382)[c] | January 1979–December 2010 | http://cfs.ncep.noaa.gov/cfsr |
| CFSv2 | NOAA/NCEP (CDAS-T574) | January 2011–present | http://cfs.ncep.noaa.gov/ |

| *Reanalysis system* | *Origin* | *Period* | *Website* |
|---|---|---|---|
| COSMO Regional Reanalysis (Europe) | Hans-Ertel-Centre, German Meteorological Service (DWD) University of Bonn | 1995–present | http://reanalysis.meteo.uni-bonn.de/?Overview |
| ERA-20Ca | ECMWF[d] | January 1900–December 2010 | http://apps.ecmwf.int/datasets/ |
| ERA-40 | ECMWF | September 1957–August 2002 | http://apps.ecmwf.int/datasets/ |
| ERA-Interim | ECMWF | January 1979–present | http://apps.ecmwf.int/datasets/ |
| JRA-25/JCDAS[e] | JMA[f] and CRIEPI[g] (JRA-25) | January 1979–January 2014 | http://jra.kishou.go.jp/JRA-25/index_en.html |
| JRA-55b[h] | JMA | January 1958–present | http://jra.kishou.go.jp/JRA-55/index_en.html |
| MERRA[i] | NASA GMAO[j] | January 1979–February 2016 | https://gmao.gsfc.nasa.gov/merra/ |
| MERRA-2c | NASA GMAO | January 1980–present | https://gmao.gsfc.nasa.gov/research/merra/ |
| NCEP-DOE[k] R2 | NOAA/NCEP and the DOE AMIP[l] -II project(R2) | January 1979–present | https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html |
| NCEP-NCAR[m] R1, | NOAA/NCEP and NCAR(R1) | January 1948–present | https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html |
| NCEP NARR[n] | NCEP | 1979–near present | http://www.emc.ncep.noaa.gov/mmb/rreanl/ |
| NOAA-CIRES[o] 20CR v2d, | NOAA and the University of Colorado CIRES(20CR) | November 1869–December 2012 | https://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html |

[a] Arctic System Reanalysis (ASR)
[b] Climate Forecast System Reanalysis (CFSR)
[c] National Oceanic and Atmospheric Administration (NOAA), National Center for Environmental Prediction (NCEP), Climate Data Assimilation System (CDAS)
[d] European Centre for Medium-Range Weather Forecasts (ECMWF)
[e] Japanese 25-year Reanalysis (JRA-25) project/Japan Meteorological Agency (JMA) Climate Data Assimilation System (JCDAS)
[f] Japan Meteorological Agency (JMA)
[g] Central Research Institute of Electric Power Industry (CRIEPI)
[h] Japanese 55-year Reanalysis (JRA-55)
[i] Modern-era Retrospective Analysis for Research and Applications (MERRA)
[j] National Aeronautics and Space Administration (NASA), Global Modeling and Assimilation Office (GMAO)
[k] Department of Energy (DOE)
[l] Atmospheric Model Intercomparison Project (AMIP)
[m] National Center for Atmospheric Research (NCAR)
[n] North American Regional Reanalysis (NARR)
[o] Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado

## 6.3 CLIMATE MODELS AND CLIMATE OUTLOOKS

The climate system, its behaviour, its components and their interactions, and its future possible evolution and changes can be simulated and studied using climate models. Climate models are being developed with increasing resolutions and complexities particularly as our understanding of the climate system has grown and larger and faster computers have become available. Models can be used for a range of applications including climate predictions and climate projections, after their representations of some elements and processes of the climate system are investigated and are shown to be skilful.

Climate outlooks providing the expected average or accumulated value of a climate element, typically over a period of months and several years, are derived from the analysis and interpretation of observations and climate model outputs.

Among the simplest models are those based on empirical or statistical relationships between large-scale fields (e.g. sea-surface temperatures, winds) and surface weather variables (e.g. rainfall, temperature, wind speed). The most complex models, global climate models (GCMs, see 6.3.4), analyse and couple the climate system for the entire globe and are used to model climate into the future explicitly or under certain assumptions. Regional climate models concentrate on representing the climate on smaller space scales over a limited area. The use of global and/or regional climate models is referred to as a dynamical approach. In some cases it is a hybrid approach (combining the statistical and dynamical methods) that is used. This method uses relationships between GCM outputs and historical data. Since numerical predictions from GCMs need vast computer resources, there are only a small number of climate centres that perform operational numerical climate predictions. In general all these models are used extensively to the benefit of climate outlooks.

### 6.3.1 **Climate outlook products**

Climate outlooks are forecasts of the values of climate elements averaged or accumulated over timescales of about one month to several years. Seasonal forecasts are more commonly used and can be generally issued with monthly frequency. Alternatively, some forecasts are issued only for specific seasons or at other pre-defined intervals. The climate elements typically forecast are average surface air temperature and total precipitation. Increasingly other parameters such as the number of days with precipitation, snowfall, the frequency of tropical cyclones and the onset and cessation of monsoon seasons are also forecast in some centres.

Forecasts of tropical sea-surface temperatures can also be regarded as climate forecast products given the importance of sea-surface temperature (SST) forcing in tropical regions. In particular, ENSO forecasts are in great demand and are largely disseminated by most of the Global Producing Centres for Long-range Forecasts (GPCLRFs) and other international institutions. Ultimately, the forecast element, frequency and lead time can vary depending on the characteristics of the climate of a particular country or region and, importantly, should be related to the users' need.

The World Meteorological Organization collaborates with its Members to produce global seasonal outlooks such as the El Niño/La Niña Update and the Global Seasonal Climate Update (GSCU). The El Niño/La Niña Update is a consensus-based product that uses inputs from a worldwide network of forecasting centres and provides an indication of the phase and strength of the ENSO. The GSCU summarizes the current status and the expected future behaviour of seasonal climate, in terms of major general circulation features and large-scale oceanic anomalies around the globe, as well as their likely impacts on continental-scale temperature and precipitation patterns. The GSCU is prepared through a collaborative effort of the WMO Lead Centre for Long-range Forecast Multi-model Ensemble prediction (LC-LRFMME), GPCLRFs and the National Oceanic and Atmospheric Administration (NOAA) among others. Both updates should be considered as complementary to more detailed regional and national seasonal climate outlooks, such as those produced by Regional Climate Outlook Forums and NMHSs.

Operational climate forecasts are produced by NMHSs, Regional Climate Centres (RCCs), GPCLRFs and other international institutions. Ideally these institutions provide their forecasts in accordance with the *Manual on the Global Data-processing and Forecasting System* (WMO-No. 485). The Global Data-processing and Forecasting System comprises World Meteorological Centres, Regional Specialized Meteorological Centres and NMHSs, which carry out functions at the global, regional and national levels (see Figure 6.1). The System represents the interactions that can be anticipated when providing a forecast: climate outlook products from a global climate model are downscaled to the regional level and further to the national level. Several NMHSs make forecasts for their own countries based on products from GPCLRFs, RCCs and other international institution.

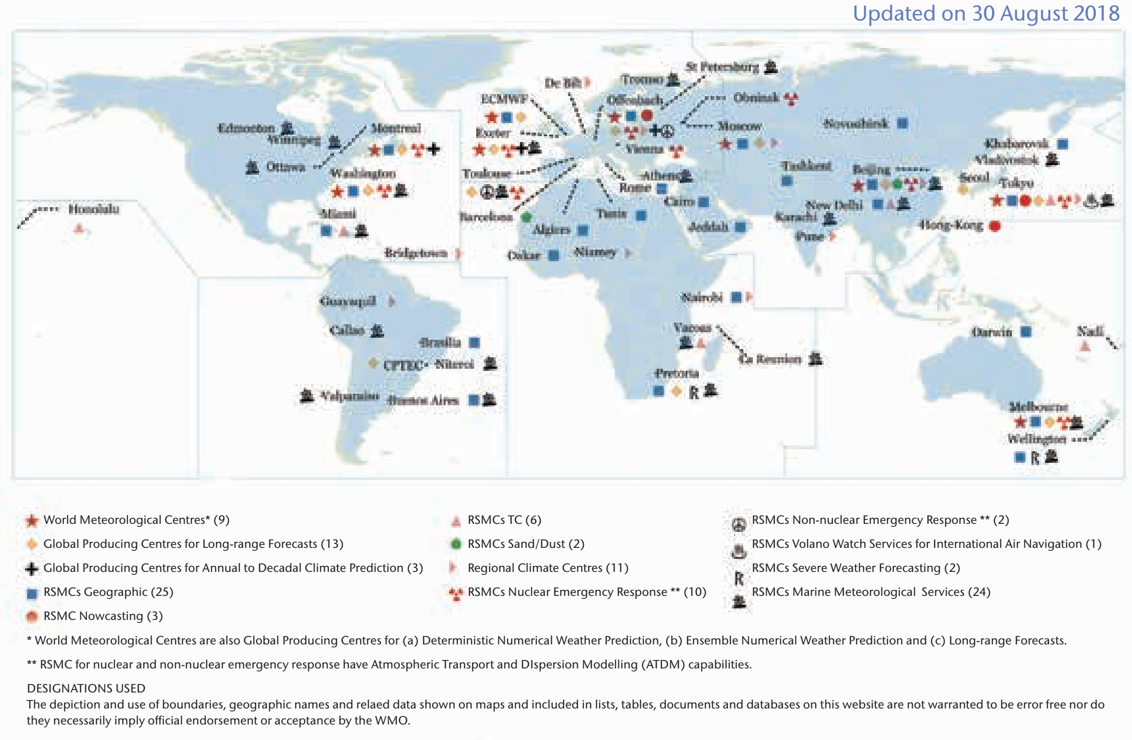



**Figure 6.1. WMO Global Data-processing and Forecasting System Centres**

Forecast products can be provided directly to specific users, and climate services are often required to interpret the meaning of the forecasts to the user (see Chapter 7). In parallel, Regional Climate Outlook Forums (RCOFs) cooperatively produce consensus-based climate outlooks using input (climate predictions) from national, regional and international climate experts. By bringing together countries with similar climatological characteristics, the Forums ensure consistency in access to, and interpretation of, climate information. Additionally, through interaction with users in the key economic sectors of each region, extension agencies and policymakers, the Forums assess the likely implications of the outlooks for the most pertinent socio-economic sectors in a given region, and explore the ways these outlooks could be used by them. News media and the Internet are other common means of disseminating forecasts to the public.

The World Climate Research Programme (WCRP) has a number of projects that aim to support climate outlooks on varying timescales. For example, the Subseasonal-to-Seasonal (S2S) Prediction Project, launched in 2013, explores sources of predictability on the sub-seasonal to seasonal timescale. The Decadal Climate Prediction Project (DCPP), which uses GCMs to explore multiyear and decadal predictions, has been included in the Coupled Model Intercomparison Project (CMIP) Experiment Design. One of its components involves the ongoing production, analysis and dissemination of experimental, quasi-real-time multi-model forecasts as a basis for potential operational forecast production. These activities support the WCRP Grand Challenge on Near-Term Climate Prediction, which has been designed to fill an important gap in the provision of seamless climate information that is bounded by seasonal-to-inter-annual climate predictions on the one hand, and multi-decadal and longer-term climate change projections on the other. These efforts strive to provide a seamless climate service, as recommended by the Global Framework for Climate Services (GFCS).

Uncertainty is unavoidable in climate forecasting due to the chaotic nature of the atmosphere, the challenges of observational data (e.g. quality, length, density of network, access to metadata and insufficient samples of extreme events) and the approximations of the forecast models. Uncertainties can also be related to lead time, the variable of interest, and temporal and spatial scales. In this context, deterministic forecasts are generally unreliable. However, forecasts with uncertainty estimates, such as  probabilistic forecasts, are essential. The forecast elements are generally represented by categories such as above normal, near normal and below normal. The probabilistic forecast gives the chance of occurrence of a category. Another representation can be achieved with the provision of the most likely category. Probabilistic forecasts, however, are more difficult to apply; users need to become familiar with the merits and limitations of probabilistic forecasts and also with methods of cost–benefit analysis. When forecasts of elements are presented as numerical quantities, the forecast uncertainty can be expressed with confidence limits or by attaching verification statistics of past forecasts. The climate service should also consider the results of past forecast verification experiments to guide the use of probabilistic forecasts.

### 6.3.2 **Climate predictions and projections**

Generally speaking, a prediction is an output of a model that computes the evolution of targeted parameters from initial conditions up to the final state at seasonal, annual or decadal timescales. A prediction assumes that factors beyond those explicitly or implicitly included in the prediction model will not have a significant influence on what is to happen. In this sense, a prediction is most influenced by the current conditions that are known through observations (initial conditions) and assumptions about the physical processes that will determine future evolutions. For example, a climate prediction that a major ENSO event will develop over the next few months is mostly determined by the state of the climate system as observed currently and in the recent past. The small changes that may occur over the next few months and other factors that are potentially influential on longer timescales, such as human activities (boundary conditions), are likely to be less important to the climate forecast.

Predictions can be formulated in either a deterministic or probabilistic approach. The deterministic forecast presents a single value attached to a specific period. The probabilistic forecast gives a range of possible values over a specific period. A prediction is made probabilistic by accounting for various types of uncertainties. A probabilistic forecast is made with several individual forecasts from a climate model starting with slightly different initial conditions (both atmospheric and oceanic) and generating a set (or ensemble) of forecasts. Ideally the sampling related to model uncertainty is also assessed using several models (multi-model ensembles). The *Guidelines on Ensemble Prediction Systems and Forecasting* (WMO-No. 1091) provides useful guidance on this approach. A climate prediction is a statement about the likelihood that something will occur, irrespective of human intervention. This is important for decision-makers, as their actions would have no impact on weather or climate events.

A climate projection is usually a statement about the likelihood that something will happen several decades to centuries in the future, if certain influential conditions develop. In contrast to a prediction, a projection specifically allows for significant changes in the set of boundary conditions, such as an increase in greenhouse gases, which might influence the future climate. As a result, what emerges are conditional expectations (if this happens, then that is what is expected). For projections extending well into the future, scenarios are developed (see 6.3.3) of what could happen given various assumptions and judgments.

As with climate prediction, the uncertainty inherent in projections should be assessed using the different possible scenarios and several models to sample also the model uncertainty. For decision-makers, they indicate the likely outcomes resulting in part from the adoption of specific policy-driven actions.

### 6.3.3 **Climate scenarios**

Global climate models are largely used in the generation of climate scenarios. A climate scenario refers to a plausible future climate constructed for investigating the potential consequences of human-induced climate change, but still representing future conditions that account for natural climate variability. The reports of the Intergovernmental Panel on Climate Change (IPCC) (for example, IPCC, 2013) are a good source of information about future climate scenarios on a timescale of several decades to a century. For example, the *IPCC Fifth Assessment Report* (AR5) included projections premised on a range of possible radiative forcing values in the year 2100 relative to pre-industrial values (+2.6, +4.5, +6.0 and +8.5 W/m2) (see Summary for Policymakers, in: *Climate Change 2013: The Physical Science Basis*, section E.7, Table SPM.3). These pathways are called Radiative Concentration Pathways (respectively RCP2.6, 4.5, 6 and 8.5). They are used as input to GCMs whose output and subsequent analyses provide descriptions of four possible climate futures, all of which are considered possible depending on how much greenhouse gas is emitted in the years to come. By considering how these RCPs would affect the climate using different climate models and each with its own particular climate sensitivity, the projections of climate change account for a wide range of reasonable possibilities of both societal development and climate behaviour. While the use of RCPs as input to climate models is the latest approach to obtaining possible future climates, previous inputs that have been extensively used include the scenarios described in the IPCC Special Report on Emissions Scenarios (SRES) (e.g. A1B, B2).

### 6.3.4 **Global climate models**

Global Climate Models are designed mainly for representing climate processes on a global scale. They provide the essential means to study climate variability and climate change for the past, present and future. They are based upon the physical laws governing the processes and interactions of all of the components of the climate system, expressed in the form of mathematical equations in three dimensions. The highly non-linear governing equations are solved numerically on a four-dimensional grid of the atmosphere (three space dimensions plus time). Many physical processes such as individual clouds, convection and turbulence take place on much smaller spatial and temporal scales than can be properly resolved by the grid. These processes have to be included through a simplified representation called parametrisation: this is a method used to replace small-scale, complex processes that are in the model with a simplified process.

These models first became practicable in the 1960s, and since then they have undergone rapid development and improvement. They have developed in parallel with numerical weather prediction models. Initially, GCMs were directed at coupling the atmosphere and ocean; most state-of-the-art GCMs now include representations of the cryosphere, biosphere, land surface, and atmosphere chemistry and aerosols in increasingly complex integrated models that are sometimes called earth system models.

Confidence in GCMs has increased substantially as a result of systematic model intercomparisons (e.g. CMIP), the ability shown by some models to reproduce major trends in the climate of the twentieth century and in some paleoclimates, and the improved simulation of major general circulation features related to phenomena such as ENSO. In general, over many parts of the world, GCMs provide credible climate simulations at subcontinental scales and for seasonal to decadal timescales, and are therefore considered as suitable tools to provide useful climate predictions and projections. These GCMs have formed the basis for the climate projections in IPCC assessments and contribute substantially to operational seasonal forecasting especially through the GPCLRFs community and climate outlook forums.

There is considerable interest in refining the spatial resolution of GCMs in order to simulate climate on smaller scales, where most impacts are felt and adaptive capacity exists. Smaller-scale climates are determined by an interaction of forcings and circulations on global, regional and local spatial scales and on sub-daily to multi-decadal temporal scales. The regional and local forcings are caused by complex topography and land-use characteristics, features at land–ocean interfaces, regional and local atmospheric circulations such as sea breezes and tropical storms, the distribution of aerosol particles and atmospheric gases, and the effects of lakes, snow and sea ice. The climate of an area may also be strongly influenced through teleconnection processes by forcing anomalies in distant locations. The processes are often highly non-linear, making projection and prediction difficult at global scales. As a response to these modelling challenges, Regional Climate Models (RCMs) and Statistical Downscaling Models have been developed to obtain climate information at a finer spatial resolution. Regional Climate Models are nested, limited-area, high-resolution models within a coarser global model. Statistical downscaling involves the application of statistical relationships between the larger and smaller scales that have been identified in the observed climate. Both approaches can be used to provide more relevant information at regional and local scales for applications through climate predictions and climate projections.

### 6.3.5 Downscaling: regional climate models

Global Climate Models cannot provide direct information for scales smaller than their own resolution. A process known as downscaling relates the properties of a large-scale model to smaller-scale regional and local climates. The approach can be either dynamical or statistical, or a combination of both. However prior to downscaling, it is important to assess the influence of the large scale features on the climate of the region or location of interest. Indeed, the downscaling can bring additional information only if the smaller scales are significantly forced by the larger ones. For instance, the orographic effect on a tropical island can be significantly forced by the intensity of the trade winds.

The dynamical approach involves RCMs that typically use the synoptic and larger-scale information from a GCM to simulate a regional climate. Regional Climate Models are able to provide data up to a resolution of a few kilometres. A major challenge for these models is how to relate the coarse-resolution grid cell information from the GCM through the boundaries onto the finer grid cells of the RCM. In addition, specific modelling challenges include parameterization of the convection and initial conditions at relevant scales. These challenges necessitate careful validation of the GCM and RCM before attempting to use either or both of their outputs. Since the RCM is essentially driven by the GCM, good performance of the GCM is of prime importance for the smaller-scale modelling. The approach described here is generally used for climate change studies and projects (e.g. CORDEX experiments) rather than for operational seasonal forecasting as it requires huge computing resources and the provision of coupling files from the GCM in quasi-real time for seasonal forecasting. For climate change studies, it is of particular importance to assess the model uncertainty in addition to the scenario uncertainty. In this respect, the recommendation is to use several RCMs ideally coupled with several GCMs (see CORDEX experiment design).

Statistical downscaling involves the creation of statistical relationships between the large-scale variables and regional and local variables. The large scale output from the GCM is used in the statistical model to obtain local and regional future climates. Major limitations of statistical downscaling are the availability of adequate historical observations (upon which the statistical relationships are based) and the assumption that the statistical relationships are valid for future climate regions (otherwise known as stationarity). A variety of methods can be used but, generally speaking, the more complex the method, the less robust are the results. Intercomparison projects such as the Statistical and Regional Dynamical Downscaling of Extremes for European regions (STARDEX) and CORDEX have been undertaken. Linear methods are commonly used for operational purposes. Validation of a statistical model is essential. Care must be taken in the interpretation of model skill as represented by skill scores.

Hybrid methods use information provided by a GCM or a RCM in a statistical model to adapt the model signal to the properties of the regional or local climate. They are quite popular in seasonal forecasting as they allow for correction of some bias (especially spatial bias) introduced by the modelling part and consequently are a clear source of potential improvement of the quality of the forecast. Access to the hindcast of the GCM (for Model Output Statistic) or to a set of reanalyses (for Perfect Prog method) is crucial for the calibration of such models. In contrast to the statistical methods, the interactions within the climate system are expected to be represented through the large scale information provided by the GCM. Obviously, the same comments made on the statistical methods are also relevant for the statistical layer of the hybrid methods.

All the methods are prone to uncertainties caused by lack of knowledge of the Earth system, model parameter and structure approximations, randomness and human actions. The additional uncertainty introduced by the downscaling process must be assessed even if validation and verification of the results of a downscaled model are quite difficult, especially if there is an inadequate observational base. In this respect, the use of spatialisation techniques (see 5.9.3) can help to interpolate the ground observations at the relevant spatial resolution and to provide a relevant reference dataset for validation and verification purposes.

## 6.4 EXAMPLES OF PRODUCTS AND DATA DISPLAYS

Data can be presented in a number of ways. Figures 6.2 to 6.13 as well as the figures throughout this Guide illustrate some of the many simple but effective ways of presenting information.

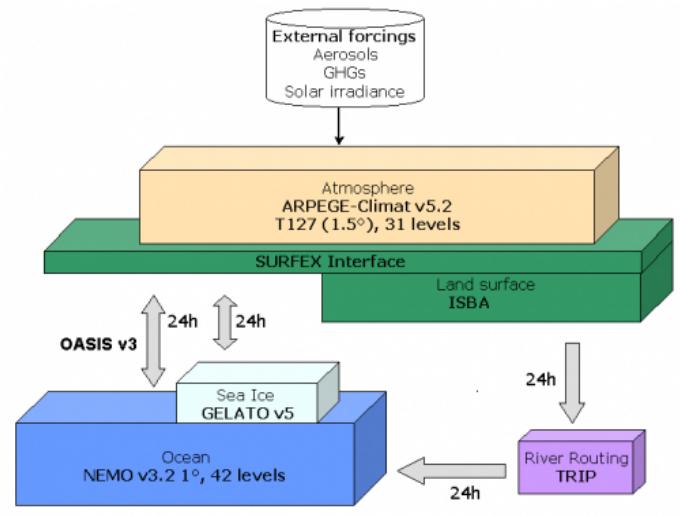



**Figure 6.2. Example of a graphic showing the Earth system components of a model designed to run climate simulations**

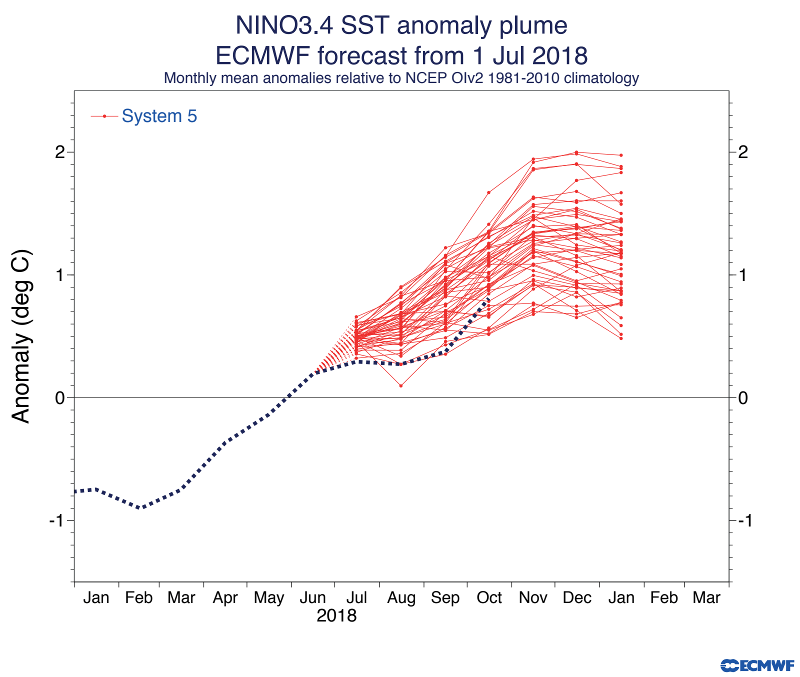



**Figure 6.3. Example of plumes representing the spread of a seasonal forecast (e.g. El Niño plumes)**

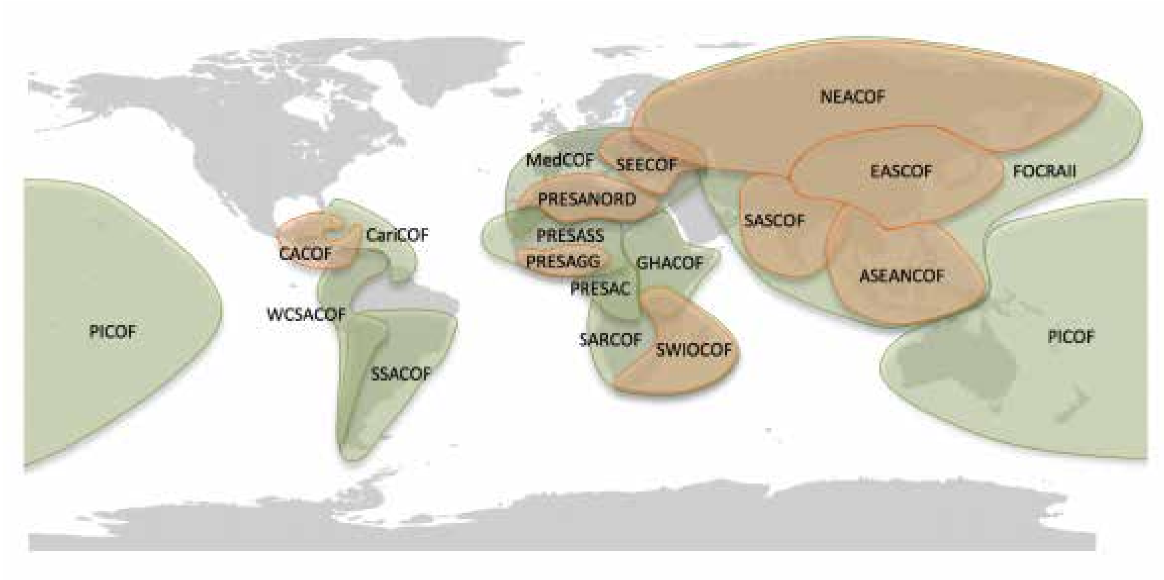



**Figure 6.4. Example of Regional Climate Outlook Forum products (consensus-based forecast)**

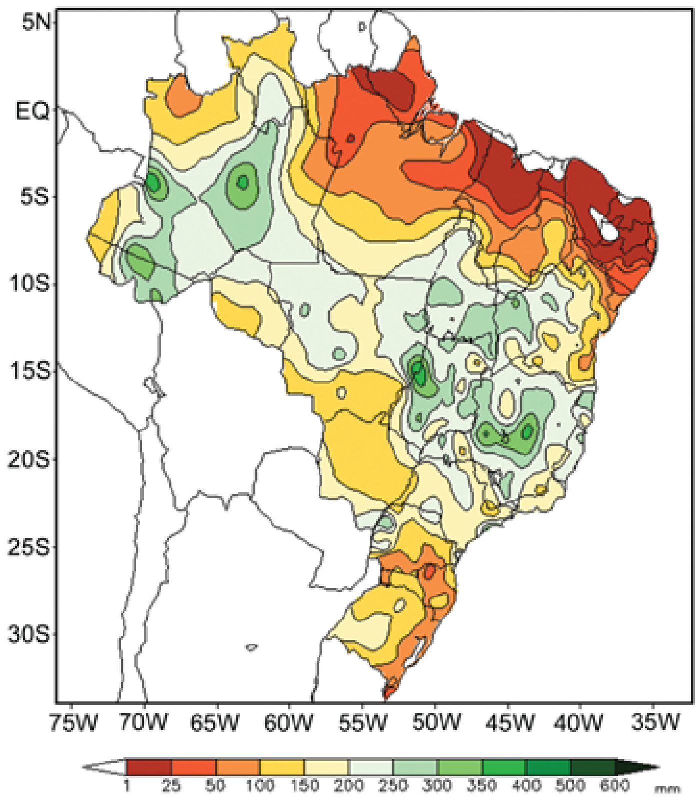



**Figure 6.5. Example of a contour map of precipitation**

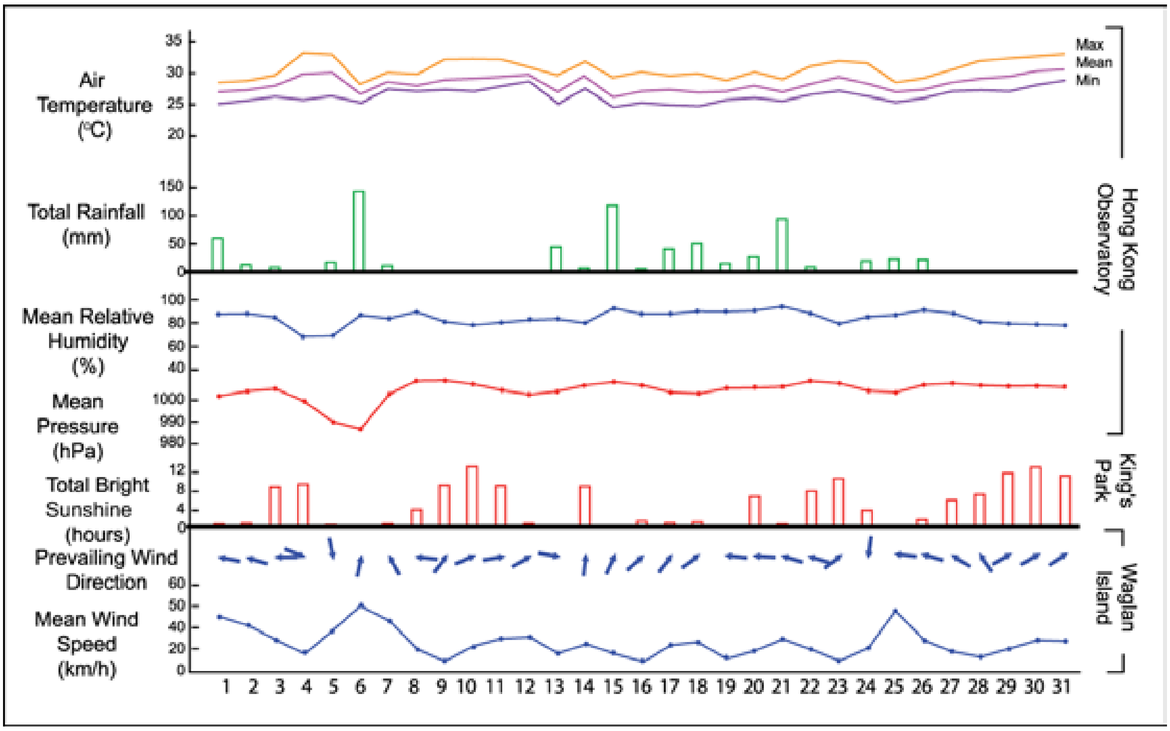
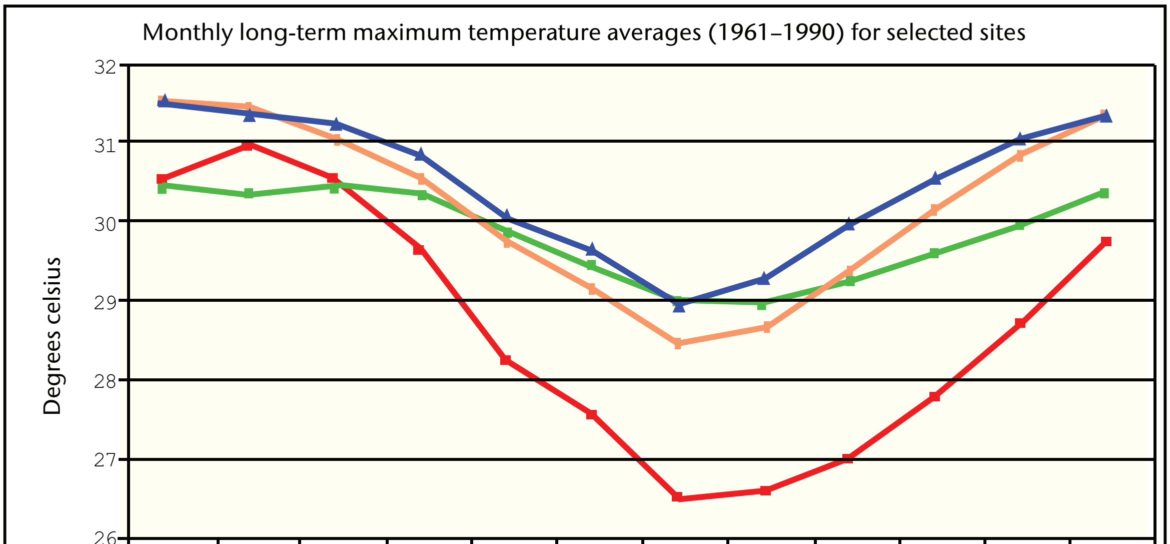



**Figure 6.6. Example of a graphical display of daily values of several elements**




| | Jan. | Feb. | March | April | May | June | July | Aug. | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nadi | 31.6 | 31.5 | 31.1 | 30.6 | 29.8 | 29.2 | 28.5 | 28.7 | 29.4 | 30.2 | 30.9 | 31.4 |
| Suva | 30.6 | 31 | 30.6 | 29.7 | 28.3 | 27.6 | 26.5 | 26.6 | 27 | 27.8 | 28.8 | 29.8 |
| Labasa | 31.6 | 31.4 | 31.3 | 30.9 | 30.1 | 29.7 | 29 | 29.3 | 30 | 30.6 | 31.1 | 31.4 |
| Rotuma | 30.5 | 30.4 | 30.5 | 30.4 | 29.9 | 29.5 | 29 | 29 | 29.3 | 29.6 | 30 | 30.4 |

**Figure 6.7. Example of a line graph and table of temperature values for multiple stations**

MELVERN LAKE, KS (145210)
Period of Record Monthly Climate Summary
**Period of Record : 5/ 1/1973 to 4/30/2000**

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Annual |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average Max. Temperature (F) | 37.3 | 43.4 | 55.3 | 65.3 | 74.1 | 82.9 | 89.1 | 87.9 | 79.6 | 68.5 | 53.6 | 42.1 | 64.9 |
| Average Min. T emperature (F) | 16.5 | 21.3 | 32.0 | 42.5 | 53.0 | 62.4 | 67.5 | 65.0 | 55.4 | 44.4 | 32.5 | 22.0 | 42.9 |
| Average Total Precipitation (in.) | 0.95 | 1.20 | 2.58 | 3.33 | 5.06 | 5.24 | 4.28 | 3.69 | 3.81 | 3.00 | 2.62 | 1.41 | 37.17 |
| Average Total SnowFall (in.) | 2.5 | 1.8 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.4 | 0.4 | 5.5 |
| Average Snow Depth (in.) | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Percent of possible observations for period of record:**

Max. Temp.: 95.9% Min. Temp.: 96.2% Precipitation: 96% Snowfall: 89.3% Snow Depth: 91%

**Figure 6.8. Example of tabular display of a monthly climate summary**

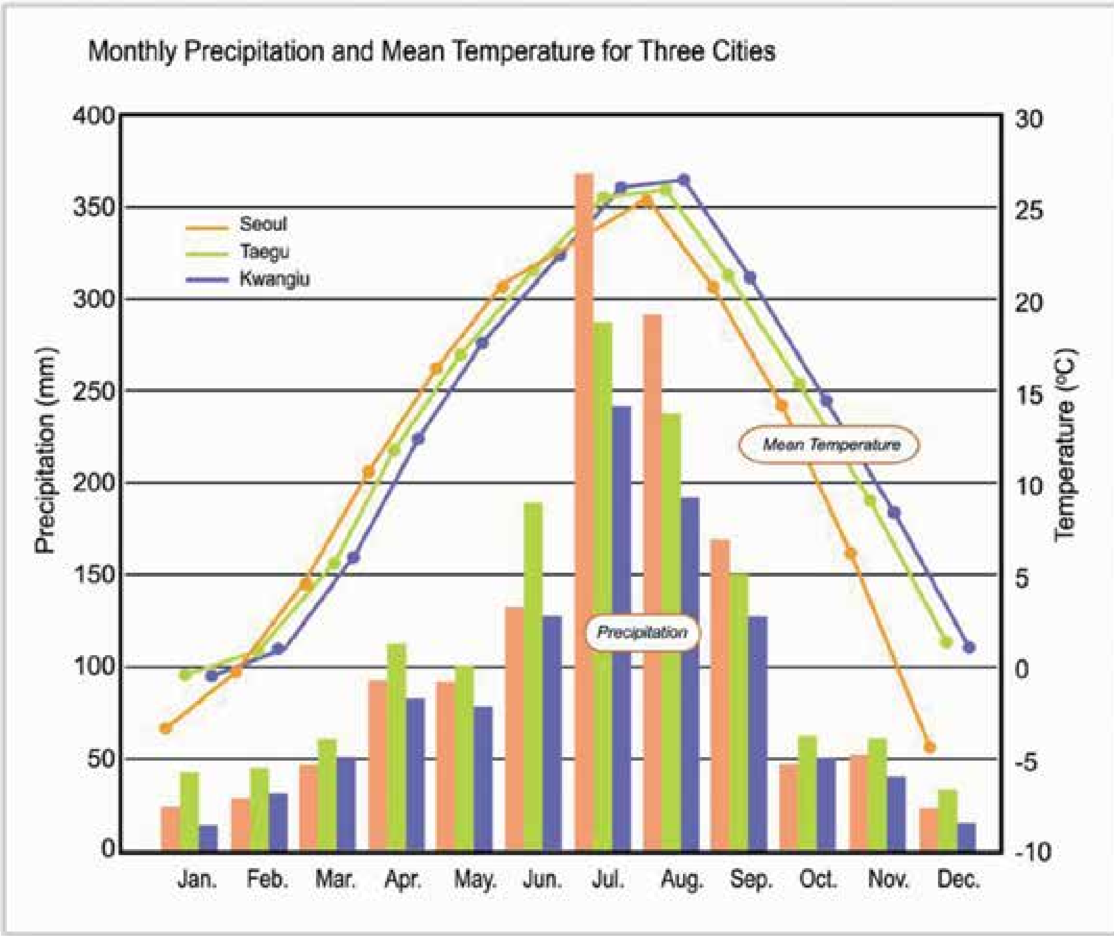



**Figure 6.9. Example of a composite line and bar chart for multiple elements and stations**

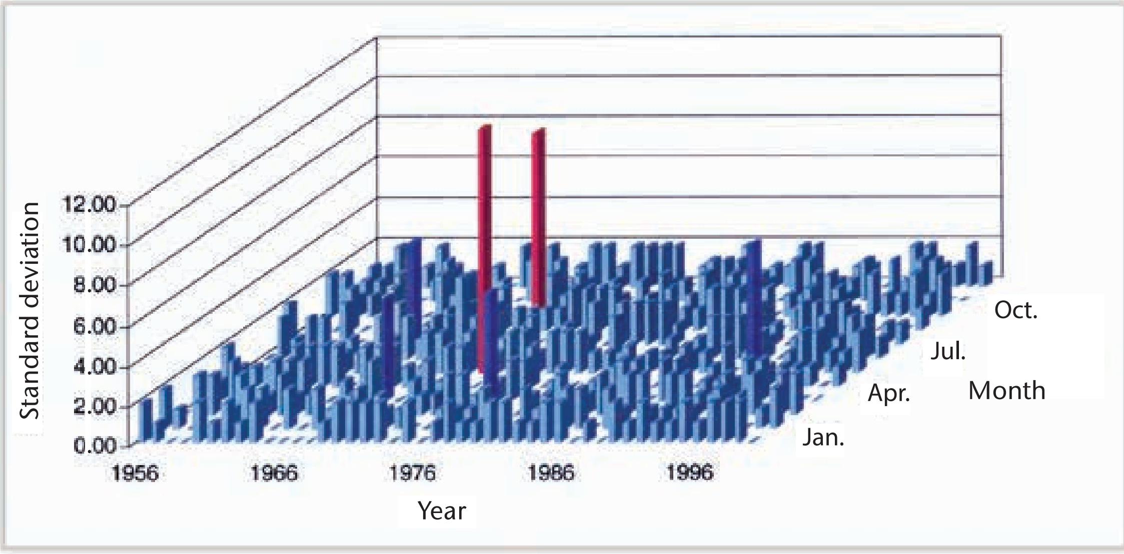



**Figure 6.10. Example of a three-dimensional display emphasizing anomalous values**

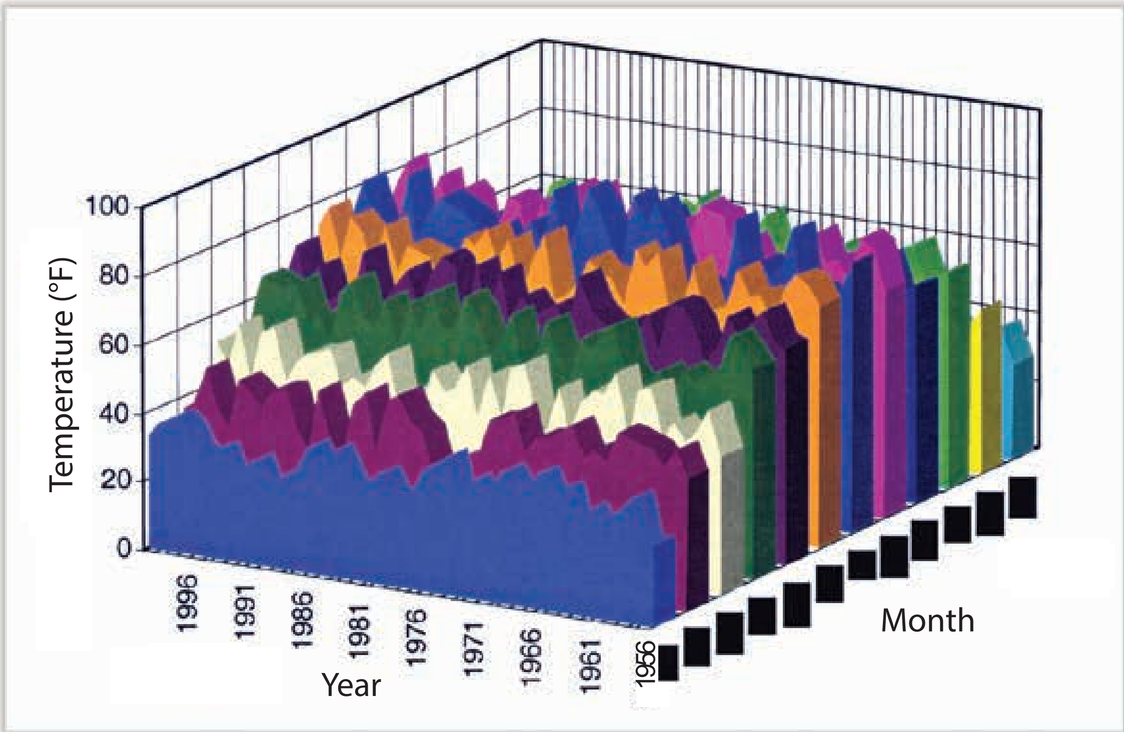



**Figure 6.11. Example of a three-dimensional display emphasizing time series of monthly values for a given year**

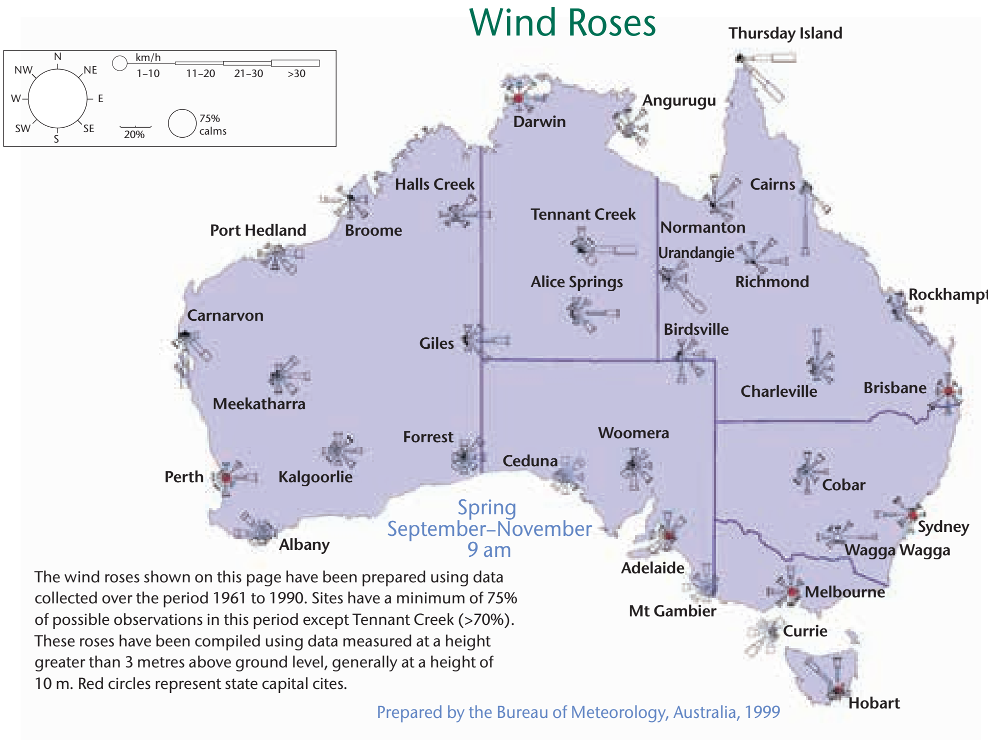



**Figure 6.12. Example of a symbolic map display of wind**

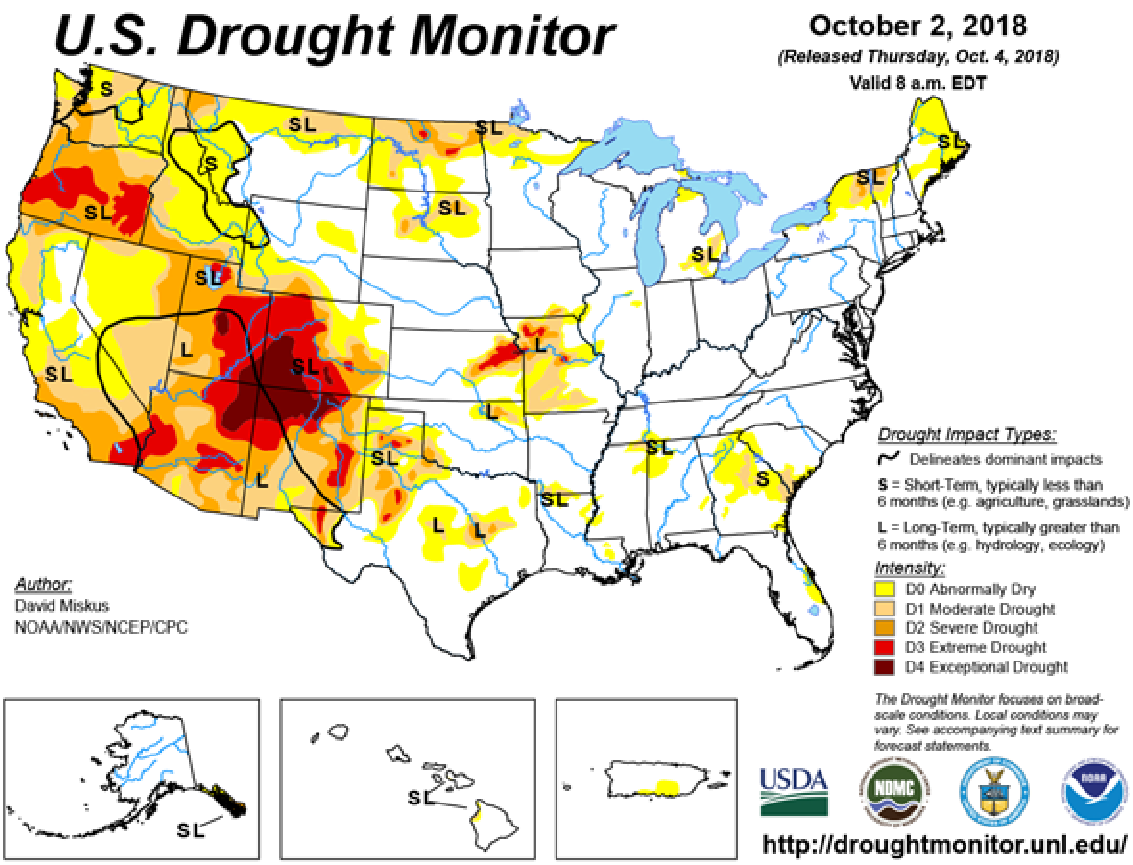



**Figure 6.13. Contour map of drought categories and impacts**

## REFERENCES


Dake, C. et al., 2004: Predictability of El Niño over the past 148 years. *Nature,* 428 (2004):733–736.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau (Jr.), Y. Zhu and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast data set. *Bulletin of the American Meteorological Society*, 94, 1553-1565.

Intergovernmental Panel on Climate Change (IPCC), 2000: *IPCC Special Report: Emissions Scenarios – Summary for Policymakers.*

———, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge, United Kingdom, and New York, United States of America, Cambridge University Press.

———, 2013: Annex III: Glossary [Planton, S. (ed.)]. In: *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, eds.). Cambridge, United Kingdom, and New York, United States of America, Cambridge University Press.

Katragkou, E., M. García-Díez, R. Vautard, S. Sobolowski, P. Zanis, G. Alexandri, R. M. Cardoso, A. Colette, J. Fernandez, A. Gobiet, K. Goergen, T. Karacostas, S. Knist, S. Mayer, P. M. M Soares, I. Pytharoulis, I. Tegoulias, A. Tsikerdekis and D. Jacob, 2015: Regional climate hindcast simulations within EURO-CORDEX: Evaluation of a WRF multi-physics ensemble. *Geoscientific Model Development*, 8, 603-618, https://www.geosci-model-dev.net/8/603/2015/.

World Meteorological Organization, 2009: *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation* (WMO/TD-No. 1500, WCDMP-No. 72). Geneva.

———, 2012: *Guidelines on Ensemble Prediction Systems and Forecasting* (WMO-No. 1091). Geneva.

_____, 2016: Regional Climate Outlook Forums, https://library.wmo.int/opac/doc_num.php?explnum_id=3191.

_____, 2017: *Manual on the Global Data-processing and Forecasting System* (WMO-No.485). Geneva.

WMO METEOTERM, http://www.wmo.int/pages/prog/lsp/meteoterm_wmo_en.html.


## FURTHER READING


American Meteorological Society, 1993: Guidelines for using color to depict meteorological information. *Bulletin of the American Meteorological Society*, 74:1709–1713.

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research*, 111:D12106.

Dunn, R.J.H., M.G. Donat and L. Alexander, 2014: Investigating uncertainties in global gridded data sets of climate extremes. *Climate of the Past*, 10 (6):2171-2199, doi:10.5194/cp-10-2171-2014.

Dunn, R.J.H., K.M. Willett, A. Ciavarella and P.A. Stott, 2017: Comparison of land surface humidity between observations and CMIP5 models. *Earth System Dynamics*, 8 (3):719-747, doi:10.5194/esd-8-719-2017.

Giorgi, F., C. Jones and G. R. Asrar, 2009: Addressing climate information needs at the regional level: The CORDEX framework. *WMO Bulletin* 58 (3):175-183.

Giorgi, F. and W.J. Gutowski (Jr.), 2015: Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources*, Volume 40:467-490, doi:10.1146/annurev-environ-102014-021217.

Gonçalves, M., P. Martinho, and C. Guedes Soares, 2016: A hindcast study on wave energy variability and trends in Le Croisic, France. In *Progress in Renewable Energies Offshore.* Proceedings of the 2nd International Conference on Renewable Energies Offshore (RENEW2016), Lisbon, Portugal, 24–26 October 2016. CRC Press.

Gutowski, W. J. (Jr.), F. Giorgi, B. Timbal, A. Frigon, D. Jacob, H.-S. Kang, K. Raghavan, B. Lee, C. Lennard, G. Nikulin, E. O'Rourke , M. Rixen, S. Solman, T. Stephenson and F. Tangang, 2016: WCRP COordinated Regional Downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6. *Geoscientific Model Development*, 9, 4087-4095.

Hargreaves, G.H., 1975: Moisture availability and crop production. *Transaction of the ASAE*, 18:980–984.

Hofstra, N., M. New and C. McSweeney, 2010: The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, 35 (5):841-858, doi:10.1007/s00382-009-0698-1.

Intergovernmental Panel on Climate Change (IPCC), 2014: *The Fifth Assessment Report (AR5),* https://ipcc.ch/report/ar5.

Jones, P.D., M. New, D.E. Parker, S. Martin and I.G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, 37:173–199.

Khambete, N.N., 1992: Agroclimatic classification for assessment of the crop potential of Karnataka. *Mausam*, 43(1):91–98.

McEvoy, D.J., J.F. Mejia and J.L. Huntington, 2014: Use of an observation network in the Great Basin to evaluate gridded climate data. *Journal of Hydrometeorology*, 15 (5):1913-1931, doi:10.1175/JHM-D-14-0015.1.

Peterson, T.C., 2005: Climate change indices. *WMO Bulletin*, 54:83–86.

Peterson, T.C. and M.J. Manton, 2008: Monitoring changes in climate extremes: A tale of international cooperation. *Bulletin of the American Meteorological Society*, 89:1266–1271.

Randall, D.A., R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi and K.E. Taylor, 2007: Climate models and their evaluation. In: *Climate Change 2007: The Physical Science Basis.* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller, eds.). Cambridge, United Kingdom, and New York, United States of America, Cambridge University Press.

Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett, 2006: Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, 19:446–469.

Sarker, R.P. and B.C. Biswas, 1988: A new approach to agroclimatic classification to find out crop potential. *Mausam*, 39(4):343–358.

Stouffer, R. J., V. Eyring, G.A. Meehl, S. Bony, C. Senior, B. Stevens and K.E. Taylor, 2017: CMIP5 scientific gaps and recommendations for CMIP6. *Bulletin of the American Meteorological Society*, 98 (1):95–105, doi:10.1175/BAMS-D-15-00013.1.

Taylor, K.E., R. J. Stouffer and G.A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93 (4):485-498.

Tufte, E.R., 1990: *Envisioning Information*. Cheshire, Connecticut, Graphic Press.

Van Vuuren, D.P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G.C. Hurtt, Kram, T. Kram, V. Krey, J.F. Lamarque and T. Masui, 2011: The representative concentration pathways: An overview. *Climatic Change*, 109, 5–31, doi:10.1007/s10584-011-0148-z.

Von Storch, H. and Navarra, A. (eds.), 2013: *Analysis of climate variability: Applications of statistical techniques*. Proceedings of an autumn school organized by the Commission of the European Community, Elba (Italy), 30 October–6 November 1993. Springer Science & Business Media.

Wikipedia, 2017, https://en.wikipedia.org/wiki/Backtesting.

# CHAPTER 7. SERVICE DELIVERY

## 7.1 INTRODUCTION

Climate services provide climate information, products or activities that facilitate decision-making by individuals and organizations. Service is defined as the delivered product and the activities associated with the people, process and information technology required to deliver it, or as an activity carried out (for example, advice and interpretation of climate information) to meet the needs of the user or that can be carried out by a user (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No. 1129), 4.1 and 4.2). Ideally, a service should be based on an understanding of user's requirements; it should provide information, products and advice that are tailored to the user, in terms of timing, format, or content, and should maintain a dialogue with the user.

To be effective, services should be:

- Credible: for the user to confidently use them in decision-making;
- Available and timely: ready when and where required by the user;
- Dependable and reliable: consistently delivered on time and according to the required user specification;
- Usable: presented in user-specific formats so that the client can make full use of them;
- Useful: capable of responding adequately to user needs;
- Expandable: applicable to different kinds of service;
- Sustainable: affordable and consistent over time;
- Responsive and flexible: adaptable to evolving user needs;
- Authentic: guaranteed to be accepted by stakeholders in a given decision context.

Such services involve high-quality data from national and international databases, as well as maps, risk and vulnerability analyses, assessments, and long-term projections and scenarios. Depending on the user's needs, these data and information products may be combined with non-meteorological data, such as agricultural production, health trends, population distributions in high-risk areas, road and infrastructure maps for the delivery of goods, and other socio-economic variables (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No. 1129), 4.2). Service delivery is a continuous, cyclic process for developing and delivering user-focused services (see Figure 7.1). It comprises four stages:

**Stage 1**: Identifying users and understanding their needs, as well as understanding the role of climate information in different sectors;

**Stage 2**: Involving users, providers, suppliers and partners in designing and developing services and ensuring that user needs are met;

**Stage 3**: Producing and disseminating data, products and information (i.e., services) that are fit for purpose and relevant to user needs;

**Stage 4**: Collecting user feedback and developing performance metrics to continuously evaluate and improve products and services.

The Global Framework for Climate Services (GFCS) was established to provide a credible, integrative and unique platform for guiding and supporting activities implemented within climate-sensitive areas (see the *Implementation Plan of the Global Framework for Climate Services*, 1.3). It aims to enable society to manage better the risks and opportunities arising from climate variability and change. The Framework puts a strong emphasis on user involvement and capacity development, and the engagement of all partners in this concerted effort is designed to maximise benefits for all users.

## 7.2 USERS AND USES OF CLIMATOLOGICAL INFORMATION

The users of climate services are many and varied, ranging from schoolchildren to global policymakers. They include diverse groups such as the media and public information personnel, farmers, the armed forces, government departments, business and industry personnel, water and energy managers, consumers, tourists, legal professionals, health officials, humanitarian and relief organizations, and meteorological services. Their needs may range from a simple interest in the weather and climate to a school project, the design of a building, agricultural operations, water management, the operation of an air-conditioning system or a large dam, the planning of energy production and distribution, or the preparation for and response to food or water shortages.

The current interest in climate change and its consequences has brought about an additional need for climate information. In the past, climate products were mostly limited to information about the physical environment of the atmosphere near the Earth's surface. Today, users may require information about many aspects of the broader climate and Earth system (solid earth, air, sea, biosphere and land surface, and ice). Data related to climate are now used to describe, represent and predict both the behaviour of the whole climate system (including the impact of human activity on climate) and the relationship between climate and other aspects of the natural world and human society. There is, therefore, an increasing need for monitoring and describing the climate in detail, on both spatial and temporal scales, while placing current events in a range of historical perspectives. High-quality baseline climate datasets are being compiled for which complex statistical quality assurance techniques have been devised. During the compilation of such datasets, procedures are put in place to detect and, if possible, correct data for outliers, and to account for inhomogeneities such as changes in instrumentation or location of observing stations. Researchers and other expert users of climate products are also keenly interested in metadata to help interpret and analyse data homogeneity and quality.

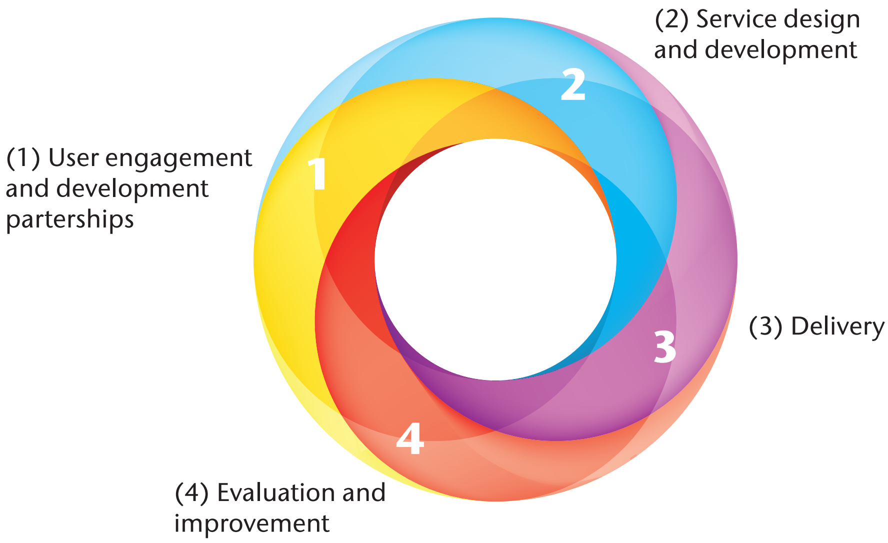



**Figure 7.1. Four stages of a continuous, cyclic process for developing and delivering services**

The uses of climatological information can be classified into two broad categories, strategic and tactical. Strategic uses refer to products that aid in the general long-term planning and design of projects and policies. Adapting to climate change is becoming a routine and necessary component of planning at all levels. In the context of climate change adaptation National Hydrological and Meteorological Services (NMHSs) are critical players in national development planning within almost all sectors (see *Climate Services for Supporting Climate Change Adaptation* (WMO-No.1170)). The types of information that are usually required for strategic uses are probability analyses and risk assessments of meteorological events for design specifications and regulations, summaries of historical conditions as background information about past climate conditions, and climate scenarios as indicators of future expectations. An example of strategic use is an analysis of climate data for designing a dam. Tactical uses refer to products and data that aid in solving short-term, specific, immediate problems. Typical information provided for tactical uses includes copies of official observations of the occurrence of a meteorological event, summaries of historical data, and the placement of an event into historical context. An example of tactical use is the analysis of recent observational data to assist in managing the use of water during a drought. Sometimes uses can be strategical and tactical at once. For example, the calculation of probabilities of the speed and direction of movement of tropical storms from historical storm data is a strategic use. But the same information, when used in forecasting the movement of a current storm, is a tactical use.

The use of climate information may include, but is not limited to:

- Monitoring of specific activities that are driven by meteorological conditions (for example, fuel consumption for heating and cooling, air pollution levels crossing thresholds, variability of sales of goods, drought-related inflation in commodity markets);
- Prediction of the behaviour of sectoral systems that react to meteorological events with a known response time, and in which the knowledge of recent past weather and climate allows some forecasting of sectoral impacts (for example, energy production and consumption, anticipation of restocking of goods, crop production, plant diseases, heat-health warning systems, food security, water supply and demand);
- Certification for insurance or other purposes that a meteorological event such as a thunderstorm, high winds, frost or drought occurred;
- Monitoring to identify the character of a given event or period, especially deviations from normal (for example, intensity of extreme rainfall or dryness);
- Design of equipment for which the knowledge of local climatology is critical to effectiveness and efficiency (such as civil engineering works, air conditioning systems, and irrigation and drainage networks);
- Impact studies, such as knowledge of initial conditions in order to assess the consequences of installing a power plant or other industrial enterprise that could affect air quality;
- Study of the influence of meteorological conditions on economic sectors, such as public transport and tourism;
- Planning and risk management for providing community services to society, such as water and energy supply, and emergency preparedness and response;
- Supporting climate change adaptation.

To promote the use of climate services, effective user engagement throughout the service development and delivery process is important (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No.1129)). Such engagement will help to increase knowledge of user needs. It also helps to gain an understanding of the impact of climate service on the protection of life and property, preservation of the environment and promotion of economic development and prosperity. This shared knowledge leads to more effective products and services that are better aligned with external demands and are fit for purpose.

The following principles embody an effective delivery of climate-related services:

- User engagement should be initiated at an early stage of the development of product and services;
- Client feedback is essential throughout the process for designing and delivering effective services;
- Sharing best practices between providers and users leads to effective and efficient service design and implementation;
- Partnerships with other international and regional organizations also engaged in service delivery are essential for maximizing the use of climate information in the decision-making process. Figure 7.2 shows the elements of importance in moving towards a service oriented culture.

## 7.3 INTERACTION WITH USERS

There can be a strong educational and communication aspect to the provision of climate services. Many people with a need for climate information have little understanding of meteorological science and related concepts. Thus, they may not know what information they really need or how best to use it. Many users may not know how best to incorporate the climate information into their own decision-making processes. Often, they simply request products that they know to be available. National Hydrological and Meteorological Services need to develop more detailed methods and tools in partnership with users. This will better integrate users into the service delivery process (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No.1129), Chapters 3 and 4).

Even the best information, issued on time, will have little impact if it does not generate the appropriate response from the users. Most of the utility of climate-related information stems


**Figure 7.2. The six elements needed for moving to a more service-oriented culture**

from the communication of this information to users, and their response based on such information. Ultimately, the utility of climate-related information depends on its social and economic impact. When available information is underutilized, value can be increased by enhancing communication between providers and users. Effective service delivery, then, is about providing products and services that are useful to users and customers. National Meteorological and Hydrological Services should not evaluate user needs in isolation, but rather in collaboration with users, providers and partners. Service provider and user need to be mutually aware of the timeline of service production and decision-making process. Having a service fit for purpose implies that an agreement has been reached, either implicitly or explicitly, among all involved, and that it takes into account some or all of the following (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No.1129), 4.1):

- Current and evolving user needs;
- Provider capabilities, including strengths and limitations;
- Types of service to be provided and how they will be provided;
- How services will be used;
- Expectations of acceptable outcomes and provider performance;
- Acceptable costs or levels of effort;
- Risks inherent in applying information to decision-making.

The climate service should ensure that expertise in the communication, interpretation and use of climate information is available. The service personnel are the link between the technical climatological aspects of the data, analyses, projections and scenarios, and users of the information who may not be technically proficient. Climate service staff should be prepared to respond to requests for data with additional information about sites, elements and instrumentation, mathematical definitions of various parameters, the ways in which observations are performed, and the science of meteorology and climatology in particular. Mechanisms to support users should be formalized and well known to users. Climate service personnel should cultivate a broad range of skills and expertise or have access to people with the necessary expertise.

Users sometimes organize meetings and activities to which climate service personnel are invited. A positive response to these invitations builds stronger relationships and gives the service personnel the opportunity to learn about users and their problems. It is important and very rewarding to be involved in users' activities as much as possible; feedback from the users usually leads to better products, new applications, and more efficient and inclusive dissemination of climate information. Continuous or frequent communication with users is essential for ensuring that existing products still meet the requirements of the users and for determining what changes need to be made to products to satisfy the users.

Customer service personnel need to be courteous and tactful and recognize the importance of professional services; user needs may be urgent for many reasons. Ideally, the climate service should have good basic communications with technological and training facilities supporting the customer service personnel. Personnel providing customer services are the people with whom the public directly interacts, the people on whom the service provider's reputation depends.

Clearly advertised contact points should be available for a variety of communication methods, together with information about service standards such as hours of operation and turn-around times. It is important to offer a variety of methods for delivering information. The service providers may have an up-to-date and powerful system backing their service, but many users may not. Modern communications and technology can now transfer and deliver data and products very quickly via the Internet, but alternative methods of delivery may be required in some countries and under certain conditions. This will ensure the sustainability of the service provided.

An NMHS often has multiple offices, and inconsistencies in standards, formats and even products can arise among offices. Frequent liaison among offices is vital; face-to-face meetings at regular intervals are recommended, and they should be scheduled at least once a year. Centralized and commonly applicable operational procedures and single reference databases help ensure consistency, as do well-written documentation, quality standards, instructions and manuals.

Charging fees for providing services can be an especially sensitive area for users and a cause of criticism and dissatisfaction if the fees are not applied consistently or if they are seen as unreasonably high. Therefore, it is important to establish a clear and transparent service pricing policy with instructions for implementing the policy. This may involve a formal policy underpinned by a set of principles and a series of practical charges for direct use by those providing the services operationally. The policy should be made available to everyone involved in providing services, as well as to the users.

Service delivery does not stop once the product or service has been delivered. User outreach and engagement must continue to ensure that services are received and acted upon and that full benefit is achieved by the user. National Meteorological and Hydrological Services should evaluate the end-to-end service delivery process and its outputs. The purpose of this evaluation is to identify the strengths of the service and areas for improvement in terms of effectiveness, efficiency, impact, satisfaction and value to its stakeholders, customers, users, partners and employees. More specifically, these evaluations should be (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No. 1129), 4.3):

- Specific: precisely targeted to the area being measured;
- Measurable: able to collect data that are accurate and complete;
- Actionable: easy to understand, interpret and act upon;
- Relevant: able to measure only things that are important and relevant to an organization’s goals and objectives. A common mistake is to measure everything, which is time consuming and produces meaningless results;
- Timely: carried out without delays;
- Agreed upon: agreeing on acceptable levels of performance is part of the evaluation of user needs or the fit for purpose assessment;
- Owned: owners should be clearly identified. Ideally, the owners should have the ability, influence and resources to take action to ensure that targets are met;
- Consistent: evaluation methods should not promote conflicting behaviour.

One of the ways of gauging customers’ satisfaction is to use surveys. Surveys may have several levels of formality, scope and standardization, ranging from frequent customer liaison visits or user workshops to bulk information gathering exercises using standardized surveys via e-mail, the Web or by telephone. Both formal and informal methods for gathering user feedback are appropriate and useful. Surveys may be conducted at regular intervals or following the delivery of a service. Customer satisfaction results can prove important when viewed alongside interviews highlighting differences between customer perception and technical performance.

The interaction with users should be part of a quality management framework set up by the management of a climate service on the basis of customers’ needs (see *Guidelines on Quality Management in Climate Services* (WMO-No. 1221), 2.2.1, Table 2, and Chapter 6).

## 7.4 INFORMATION AND DISSEMINATION OF PRODUCTS AND SERVICES

In the provision of climate services, it is essential to put the users first. As the needs of users evolve, the capabilities of service providers should also adapt over time. Methods of distributing products and services are subject to change, especially in the modern era of information technology, and it is important that Members remain agile and capable of responding to these changes (see *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No. 1129), 4.1 and 4.4).

Engagement between users and providers occurs at multiple levels ranging from relatively passive engagement through websites and web tools to much more active engagement and focused relationships. Based on user needs, this engagement could be just the provision of information, a dialogue-based focused delivery, or a very tailored and targeted service provision. These categories transition from a more passive to an active delivery mechanism and potentially increase the amount of time and expense involved in providing the service and its uptake.

How well a user interprets a provided service depends very much on how the information is presented and provided. Where practical to do so, the important facts should be shown visually with accompanying text used to qualify, emphasize and explain. The presentation should be logical, clear, concise and tailored to suit the user and the aims of the presentation. For example, the style used to impart climate information to a researcher will be different from that of an article for a newspaper or popular magazine.

The approach used to summarize data in a periodical climatological bulletin, intended for a wide range of technical and non-technical users, should be different from that used to prepare an interpretative report on some specific problem in applied climatology. Technical information should be presented to users with a limited knowledge of atmospheric sciences in a manner that is simple and understandable, while remaining scientifically correct. It is advisable to avoid specific scientific terminology and acronyms unless defined in detail.

More knowledgeable users will be better able to understand complex information and presentations, but some distillation of technical or scientific information is usually desirable. The climate service must, therefore, be able to anticipate, investigate and understand the requirements of government decision-makers, industrial and commercial interests and the general public. It must ensure that the provider of climate services is able to understand the issues and respond to questions and user needs with current knowledge and professional skill.

The Global Framework for Climate Services was established to provide a credible, integrative and unique platform for guiding and supporting activities implemented within climate-sensitive investment areas, notably agriculture, energy, disaster risk reduction, human health and water sectors, in support of both climate adaptation and mitigation. The structure of the GFCS is built upon five components, or pillars. Two of those pillars are relevant to service dissemination:

- The User Interface Platform: a structured means for users, climate researchers and climate information providers to interact at all levels;

- The Climate Services Information System: a mechanism through which information about climate (past, present and future) will be routinely collected, stored and processed to generate products and services that inform often complex decision-making across a wide range of climate-sensitive activities and enterprises.

The objective of the User Interface Platform is to promote effective decision-making with respect to climate considerations by making sure that the right information, is delivered, understood, and used at the right time and in the right amount. The Platform uses a wide range of methods designed to promote mutual understanding, including formally established committees, working groups, internship programmes, one-to-one discussion, workshops, conferences and inter-agency task teams. Communication, outreach and training approaches are equally wide-ranging. They include radio broadcasts, social media and public service announcements, and the use of technologies such as map interfaces, portals and information servers. In many areas of this work there are opportunities to build upon dialogues already well-established or that are growing in effectiveness, such as the Regional Climate Outlook Forums, community liaison working groups in the disaster management community and national health working groups. The proposed enhanced interaction between users and providers aims to reconcile the availability of credible climate information with the needs of users for information to support their decision-making. This mutual understanding can then frame an end-to-end climate service that may involve developing useful products. WMO Members that have already implemented a formal quality management system (QMS) are more likely to be focused on meeting user needs and to consider quality management a key aspect of service delivery.

The Climate Services Information System is the means by which research outputs and technological developments are transformed into improved operational climate information. The Climate Services Information System comprises a physical infrastructure of institutions, computer capabilities and tools, and operational practices. Together with professional human resources it develops, generates and distributes a wide range of climate information products and services usable at the global, regional and national scales.

## 7.5 MARKETING OF SERVICES AND PRODUCTS

Communication is vital in creating the positive message about the value of climate services. It increases awareness of the need for and benefit of such services and helps individuals and organizations to make decisions that are relevant to climate. Therefore, an overall communication strategy for climate service provision should have a list of objectives and identify ways of achieving them, using a wide variety of communications methods and media.

Communication is an important aspect of marketing, not simply advertising and selling of climate services. It also allows potential users to learn what services and products are available, realize the utility of the services and products, and gain an understanding of the value of the information, as discussed in the *WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No. 1129), 4.3). Marketing, public relations and advertising, and promoting and disseminating climate information are essential to the success of most climate services. Climate information is unlikely to be sufficient in its own right, but often needs to be complemented with non-climate information in order to be of social and economic benefit. The benefits of using a product must be marketed. This is because the relevance of climate information often is not apparent to those who are in a position to benefit from the products or to those who approve funds for climate programmes. The benefits and worth of using the products must be clearly demonstrated in social and economic terms. Studies are needed to assess the value of climatological services. Such studies should not be the responsibility of the climatologist alone, but rather a shared responsibility of all stakeholders. One way to show the effectiveness of services is by demonstrating the value of these to the users.

The characteristics of an effective marketing and communication programme include:

- Focusing on user needs by gaining a clear understanding of the user’s requirements and how the climate information is used;
- Capacity-building for the provider of climate service and users;
- Identifying the target market;
- Promoting the benefits of climate services and products to users;
- Developing a product or service for a specific need of the user and promoting its application to solving the user’s problems;
- Promoting the professional skills of climate service personnel;
- Deciding on methods of product accessibility or delivery and making alternatives available to the user;

- Evaluating the economics of the products and services;
- Informing users through promotion and public relations;
- Monitoring user satisfaction and assessing service performance and marketing efforts;
- Ensuring the credibility and sustainability of climate services by being transparent about the reliability and limitations of the products and services offered.

Capacity-building in all aspects of service delivery should be incorporated into relevant training events. Such activities should also include information on the users of the climate information, their decision-making processes and how they apply climatological information to such decisions (see *Annex to the Implementation Plan of the Global Framework for Climate Services – Capacity Development,* Chapter 5). Users often have limited understanding of the quality of the services they receive, or even exactly what they are receiving or the appropriateness of the information. Therefore, it is also highly recommended to train users and customers on how to derive the maximum benefit from those products and services and to ensure that they fully understand the capabilities of NMHSs.

Linkages to capacity development activities often have the biggest impact when they focus on services and service delivery and take into account the infrastructure and the human and institutional capacities required to enable the delivery of such services. National Meteorological and Hydrological Services will be able to contribute more effectively to the development plans of their countries if the services they provide are designed with the needs of the user in mind. This will help to ensure that the services are valued by the users and that the meteorological and hydrological services are sustained and improved.

## REFERENCES


World Meteorological Organization, 2014: *Implementation Plan of the Global Framework for Climate Services*. Geneva.

———, 2014: *Annex to the Implementation Plan of the Global Framework for Climatic Services – Climate Service Information System Component*. Geneva.

———, 2014: *Annex to the Implementation Plan of the Global Framework for Climate Services – Capacity Development*. Geneva.

———, 2014: *The WMO Strategy for Service Delivery and its Implementation Plan* (WMO-No.1129). Geneva.

———, 2016: *Climate Services for Supporting Climate Change Adaptation - Supplement to the Technical Guidelines for the National Adaptation Plan Process* (WMO-No. 1170). Geneva.

———, 2018: *Guidelines on Quality Management in Climate Services* (WMO-No.1221). Geneva.


## FURTHER READING


American Meteorological Society, 1993: Guidelines for using color to depict meteorological information. *Bulletin of the American Meteorological Society*, 74:1709–1713.

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett and P.D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research*, 111:D12106.

Hargreaves, G.H., 1975: Moisture availability and crop production. *Transaction of the ASAE*, 18:980–984.

Jones, P.D., M. New, D.E. Parker, S. Martin and I.G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, 37:173–199.

Khambete, N.N., 1992: Agroclimatic classification for assessment of the crop potential of Karnataka. *Mausam*, 43(1):91–98.

MacCracken, M., 2002: Do the uncertainty ranges in the IPCC and U.S. National Assessments account adequately for possibly overlooked climatic influences? *Climatic Change*, 52:13–23.

Palmer, W.C., 1965: Meteorological Drought. *Weather Bureau Research Paper No. 45*. Washington, DC, United States Department of Commerce.

Peterson, T.C., 2005: Climate change indices. *WMO Bulletin*, 54:83–86.

Peterson, T.C. and M.J. Manton, 2008: Monitoring changes in climate extremes: A tale of international cooperation. *Bulletin of the American Meteorological Society*, 89:1266–1271.

Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett, 2006: Improved analyses of changes and uncertainties in marine temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate*, 19:446–469.

Sarker, R.P. and B.C. Biswas, 1988: A new approach to agroclimatic classification to find out crop potential. *Mausam*, 39(4):343–358.

Tufte, E.R., 1990: *Envisioning Information*. Cheshire, Connecticut, Graphic Press.

United States Global Change Research Program Office (USGCRP), 2000: C*limate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change*. Washington, DC.

Wang, B. and Z. Fan, 1999: Choice of South Asian summer monsoon indices. *Bulletin of the American Meteorological Society*, 80:629–638.

Wilhelm, W.W., K. Ruwe and M.R. Schlemmer, 2000: Comparisons of three Leaf Area Index meters in a corn canopy. *Crop Science*, 40:1179–1183.


---

# ANNEX. INTERNATIONAL CLIMATE ACTIVITIES

## 1. COORDINATION OF CLIMATE ACTIVITIES

The Commission for Climatology is responsible for promoting and facilitating activities relating to climate and its relationship with human well-being, socio-economic activities, natural ecosystems and sustainable development. Specifically, it:

1. Coordinates and consolidates general requirements and standards for observations, data collection, archiving and data exchange for all components of the World Climate Programme;

2. Identifies best practices in management of climate data, including near-real-time data, proxy data, remote-sensing data, and metadata (information about data);

3. Promotes the dissemination of data, products and methods in support of research, applications, impact assessments and climate system monitoring;

4. Supports the preparation of authoritative statements on climate, including the annual statements on the state of the global climate and regular El Niño and La Niña updates;

5. Evaluates and reviews development and application of operational climate predictions;

6. Identifies priorities for studying the climates of natural and managed ecosystems and for providing climate information that can help alleviate problems arising from the influence of human activities on local and regional climate;

7. Supports capacity-building and technology transfer;

8. Promotes research and evaluation of the role of climate in key social and economic sectors in partnership with other WMO technical commissions, other United Nations agencies and relevant international and regional institutions;

9. Assesses the potential for application of seasonal prediction and other climatological services for social and economic benefit, including reduction of the risk of climate-related hazards and optimal utilization of climate as a resource;

10. Provides advice on issues relating to the access and availability of climatological data and services.

The Commission promotes and relies on a range of national, regional and global entities involved in climate-related matters. Apart from NMHSs, these entities include other United Nations agencies such as the Food and Agriculture Organization of the United Nations (FAO); the World Health Organization (WHO); the United Nations World Tourism Organization (UNWTO); the United Nations Human Settlements Programme (UN-Habitat); the United Nations Environment Programme (UNEP); the United Nations Development Programme (UNDP); and the United Nations Educational, Scientific and Cultural Organization (UNESCO). The Commission for Climatology also works extensively with non-governmental organizations such as the International Federation of Red Cross and Red Crescent Societies (IFRC), the International Science Council (ISC, formerly International Council for Science (ICSU)), research institutions, universities, professional associations, and development agencies such as the World Bank.

## 2. THE WORLD CLIMATE PROGRAMME

In the 1970s, it became obvious that greater international coordination and effort were needed to tackle the deficiencies in the understanding of climate and how to deal with the wide range of climatic influences on society and the environment, some beneficial and others detrimental. In 1974, the Executive Council of the World Meteorological Organization agreed that WMO should initiate an international programme on climate and laid the foundations of the World Climate Programme. In 1978, the Economic and Social Council of the United Nations requested that WMO devote particular attention to those aspects of the Programme that would provide prompt and effective assistance to national planners and decision-makers in formulating economic and social programmes and activities in their respective countries.

The First World Climate Conference, in February 1979, recognized that the problem of possible human influence on climate was a matter of special importance. The Conference Declaration stressed the urgent need "for the nations of the world (a) to take full advantage of man's present knowledge of climate; (b) to take steps to improve significantly that knowledge; (c) to foresee and to prevent potential man-made changes in climate that might be adverse to the well-being of humanity."

The Eighth World Meteorological Congress formally created the World Climate Programme in 1979, recognizing that the cooperation of many United Nations agencies and other international organizations was needed to support the Programme. Responsibility for the overall coordination of the Programme and for climate data and applications was assumed by WMO, while UNEP assumed responsibility for climate impact studies, and WMO together with ISC agreed to jointly implement the climate research programme. The structure of the World Climate Programme and the partners cooperating with WMO have evolved over time.

There are four main components of the World Climate Programme. The World Climate Data and Monitoring Programme (WCDMP) facilitates the effective collection and management of climate data and the monitoring of the global climate system, including the detection of climate variability and change. The World Climate Applications and Services Programme (WCASP) fosters scientific understanding of the role of climate in human activities, effective application of climate knowledge and information for the benefit of society, tailored climate services for users in various socioeconomic sectors, and the prediction of significant climate variations (both natural and as a result of human activity). The Climate Information and Prediction Services (CLIPS) project was established by WMO as an implementation arm of WCASP. The World Climate Impact Assessment and Response Strategies Programme (WCIRP), implemented by UNEP, assesses the impacts of climate variability and changes that could markedly affect economic or social activities and advises governments on these matters; it also contributes to the development of a range of socioeconomic response strategies that could be used by governments and communities. The World Climate Research Programme (WCRP), jointly sponsored by WMO, ISC and the Intergovernmental Oceanographic Commission (IOC) of UNESCO, seeks to improve the basic scientific understanding of climate processes to determine the predictability of climate, climate variability and change, and the extent of human influence on climate, and to develop the capability for climate prediction. The activities and projects of WCRP include studies of the dynamics and thermodynamics of the Earth's atmosphere, the atmosphere's interactions with the Earth's surface, and the global water cycle; climate variability and predictability; interaction of dynamical, radiative and chemical processes; cryosphere interaction with the rest of the climate system; and biogeochemical and physical interactions between the ocean and the atmosphere.

The Second World Climate Conference, held in 1990, recognized an urgent need to acquire comprehensive information on the properties and evolution of the Earth's climate system, to detect climate change, to support climatological applications for economic development and to develop climate science and predictions. The Eleventh World Meteorological Congress in 1991 decided that a global climate observing system should be established. This system would be based on the coordination and association of existing or planned operational and research programmes for observing the global environment, and on the further development of those programmes required to ensure continuity of information over decades.

In 1992, the Global Climate Observing System (GCOS) was formally established under a Memorandum of Understanding among WMO, IOC-UNESCO, UNEP and the International Council for Science (now ISC). While GCOS does not directly make observations or generate data products, it works closely with and builds upon existing and developing observing systems and provides a framework for integrating the systems of participating countries and organizations. The systems and networks included are:

- The GCOS Surface Network (GSN) and Upper-air Network (GUAN) (subsets of the WMO World Weather Watch Global Observing System);
- The WMO Global Atmosphere Watch (GAW);
- The WCRP/ Global Energy and Water Exchanges (GEWEX) Baseline Surface Radiation Network;
- The Global Ocean Observing System (GOOS), including ocean observing systems such as the Global Sea Level Observing System;
- The Global Terrestrial Observing System, including the Global Terrestrial Network for Glaciers and the Global Terrestrial Network for Permafrost;
- The World Hydrological Cycle Observing System.

The concept of coordinated and comprehensive Earth observations was promoted at the First Earth Observation Summit in 2003. At the Third Earth Observation Summit held in 2005, the Global Earth Observation System of Systems (GEOSS) was formally established to build on, and add value to, existing national, regional and international observation systems by coordinating efforts, addressing critical gaps, supporting interoperability, sharing information, reaching a common understanding of user requirements, and improving delivery of information to users. The Global Earth Observation System of Systems supports GCOS by expanding on the range of climate-related variables identified in the GCOS Implementation Plan and by assisting the parties in meeting their responsibilities under the United Nations Framework Convention on Climate Change (see 4 below).

The Third World Climate Conference, held in 2009, was organized by WMO in collaboration with UNESCO, UNEP, FAO, ICSU (now ISC) and other intergovernmental and non-governmental partners. The theme of the Conference was "Climate Prediction and Information for Decision-making" and its vision was for "An international framework for climate services that links science-based climate predictions and information with the management of climate-related risks and opportunities in support of adaptation to climate variability and change in both developed and developing countries". This Conference provided nations with the opportunity to jointly consider an appropriate Global Framework for Climate Services (GFCS) over the coming decades that would help ensure that every country and every climate-sensitive sector of society is well equipped to access and apply the growing array of climate prediction and information services made possible by recent and emerging developments in international climate science and technology.

## 3. THE CLIMATE AGENDA

In the early 1990s, development of increasingly close collaboration among climate-related programmes of a number of international organizations led to the establishment by WMO of the Coordinating Committee for the World Climate Programme. In 1995, a draft interagency document entitled "The Climate Agenda – International Climate-related Programmes: A Proposal for an Integrating Framework" was issued. The four main thrusts were new frontiers in climate science and prediction, climate services for sustainable development, dedicated observations of the climate system, and studies of climate impact assessments and response strategies to reduce vulnerability. The leaders of these thrusts were WMO, FAO and UNEP.

The Twelfth World Meteorological Congress endorsed the Climate Agenda in 1995, and established an Interagency Committee on the Climate Agenda. While the Committee provided some initial guidance, the fast pace of developments in climate-related matters led WMO, at the fifty-third session of its Executive Council (2001), to consider new mechanisms for climate coordination. At its sixty-first session in 2009, the WMO Executive Council agreed not to revitalize the Interagency Committee on the Climate Agenda, but to use the United Nations Delivering as One initiative to coordinate climate issues at the United Nations level.

## 4. INTERNATIONAL CLIMATE CHANGE PROGRAMMES

At the First World Climate Conference in 1979, the international climate community expressed concern that "continued expansion of man's activities on Earth may cause significant extended regional and even global changes of climate", and called for global cooperation to "explore the possible future course of global climate and to take this new understanding into account in planning for the future development of human society". At a climate conference in Villach, Austria, in 1985, scientists from 29 developed and developing countries concluded that increasing concentrations of greenhouse gases were expected to cause a significant warming of the global climate in the next century, as detailed in the *Report of the International Conference on the Assessment of the Role of Carbon Dioxide and of Other Greenhouse Gases in Climate Variations and Associated Impacts* (WMO-No. 661) (see Working Group III – Impact on climate change, Impact of greenhouse gases on climate, pp. 56–57). They also noted that past climate data may no longer be a reliable guide for long-term projects because of expected warming of the global climate, that climate change and sea-level rise are closely linked with other major environmental issues, that some warming appears inevitable because of past activities, and that the future rate and degree of warming could be profoundly affected by policies on emissions of greenhouse gases. Responding to these concerns, UNEP, WMO and ICSU (now ISC) established an Advisory Group on Greenhouse Gases in 1986 to ensure periodic assessments of the state of scientific knowledge on climate change and its implications.

In 1987, the Tenth World Meteorological Congress recognized the need for an objective, balanced and internationally coordinated scientific assessment of the understanding of the effects of increasing concentrations of greenhouse gases on the Earth's climate and the ways in which these changes may influence social and economic patterns. The United Nations Environment Programme and WMO agreed on an intergovernmental mechanism to provide scientific assessments of climate change, and in 1988 the WMO Executive Council, with support from UNEP, established the Intergovernmental Panel on Climate Change (IPCC) to consider the need for:

- Identification of uncertainties and gaps in our present knowledge with regard to climate change and its potential impacts, and preparation of a plan of action over the short term to fill these gaps;
- Identification of information needed to evaluate policy implications of climate change and response strategies;
- Review of current and planned national and international policies related to greenhouse gases;
- Scientific and environmental assessments of all aspects of the greenhouse gas issue and transfer of this and other relevant information to governments and intergovernmental organizations, so that such assessments can be taken into account in policies on social and economic development and environmental programmes.

In November 1988, the IPCC established working groups to prepare assessment reports on the available scientific information on climate change (Working Group I), on environmental, social and economic impacts of climate change (Working Group II), and on formulation of response strategies (Working Group III). The forty-third session of the United Nations General Assembly, in 1988, endorsed the action of WMO and UNEP to establish the IPCC and requested:

> ”a comprehensive review and recommendations with respect to: (a) The state of knowledge of the science of climate and climatic change; (b) Programmes and studies on the social and economic impact of climate change, including global warming; (c) Possible response strategies to delay, limit, or mitigate the impacts of adverse climate change”.

The IPCC adopted its First Assessment Report on 30 August 1990. Its findings, coupled with the outcomes of the Second World Climate Conference that same year, spurred governments to create the United Nations Framework Convention on Climate Change in March 1994. The IPCC Second Assessment Report was completed in late 1995, the third in 2001, the fourth in 2007 and the fifth in 2014. The reports issued by the IPCC are frequently used to inform decisions made under the Framework Convention. They also played a major role in the negotiations leading to the Kyoto Protocol, an international treaty, which came into force in February 2005, that builds on the Framework Convention and sets legally binding targets and timetables for reducing the greenhouse gas emissions of industrialized countries. The IPCC reports, as the sources of the best available science, guide the implementation of the Paris Agreement. Based on the scientific findings of the IPCC's Fifth Assessment Report, the Paris Agreement was a major achievement of the Conference of the Parties at its twenty-first session in 2015. The Agreement’s objective is to limit the rise in global temperature to below 2 °C and to aim for 1.5 °C. The IPCC Assessment Reports, drawing on the work of hundreds of scientists from all over the world, enable policymakers at all levels of government to take sound, evidence-based decisions. The award of the 2007 Nobel Peace Prize to the IPCC, along with Mr Al Gore, Former Vice-President of the United States of America, is testimony to the remarkable success of the IPCC process in informing policymakers as well as the general public on the scientific underpinnings of the climate change issue.

## REFERENCES


Intergovernmental Panel on Climate Change (IPCC), 2014: *The Fifth Assessment Report (AR5)*, https://ipcc.ch/report/ar5.

United Nations Framework Convention on Climate Change, 2017: The Paris Agreement, http://unfccc.int/paris_agreement/items/9485.php.

World Meteorological Organization, 1986: *Report of the International Conference on the Assessment of the Role of Carbon Dioxide and of Other Greenhouse Gases in Climate Variations and Associated Impacts* (Villach, Austria, 9–15 October 1985) (WMO-No. 661). Geneva.


## FURTHER READING


Group on Earth Observations (GEO), 2005: *Global Earth Observation System of Systems (GEOSS): 10-year Implementation Plan*. Reference Document GEO 1000R/ESA SP-1284. Noordwijk, European Space Agency Publications Division, ESTEC.

———, 2007: *GEO 2007–2009 Work Plan. Toward Convergence*. Geneva.

United Nations System Chief Executives Board for Coordination, 2008: *Acting on Climate Change: The UN System Delivering as One*. New York.

World Climate Research Programme, 2009: *The World Climate Research Programme Achievements: Scientific Knowledge for Climate Adaptation, Mitigation and Risk Management* (WMO/TD-No. 1499). Geneva.

———, 2009: *WCRP Implementation Plan 2010-2015* (WMO/TD-No. 1503). Geneva.

———, 2013: *The World Climate Research Programme Accomplishment Report*, https://www.wcrp-climate.org/images/documents/reports_flyers/WCRP_report03_2012.pdf.

For more information, please contact:

**World Meteorological Organization**

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

**Communication and Public Affairs Office**

Tel.: +41 (0) 22 730 83 14/15 – Fax: +41 (0) 22 730 80 27

Email: cpa@wmo.int

**public.wmo.int**

JN 181197