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 in 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 maintain *awareness 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 among characteristics of frequency distributions both. These comparisons answer common questions such as:
Are average temperatures taken over a specific time interval at different locations the same?

(a) Is the variability of precipitation the same at different locations?
(b) Is the diurnal temperature range at a location changing over time, and if so, how?
(c) 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 realization of the physical system of climate and weather, and, further, 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 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 not homogeneous nor are they 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 *(Guide to the GCOS Surface and Upper-Air Networks: GSN AND GUAN*, WMO/TD-No. 1106). Unfortunately, most long-term climatological datasets have been affected by a number of factors not related to the broader-scale climate. These include, among other things, 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, a change in instrument or site), while others may cause gradual biases (for example, 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 of special concern to 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 then 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 interannual 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, like temperature) or by calculating ratios or log ratios (for elements measured on a proportional scale, like 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 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 precipitation in a highly variable temperature climate than in a less variable temperature climate). 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 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 two series. Since variations may occur naturally, it is recommended that the apparent changes of the slope occur 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

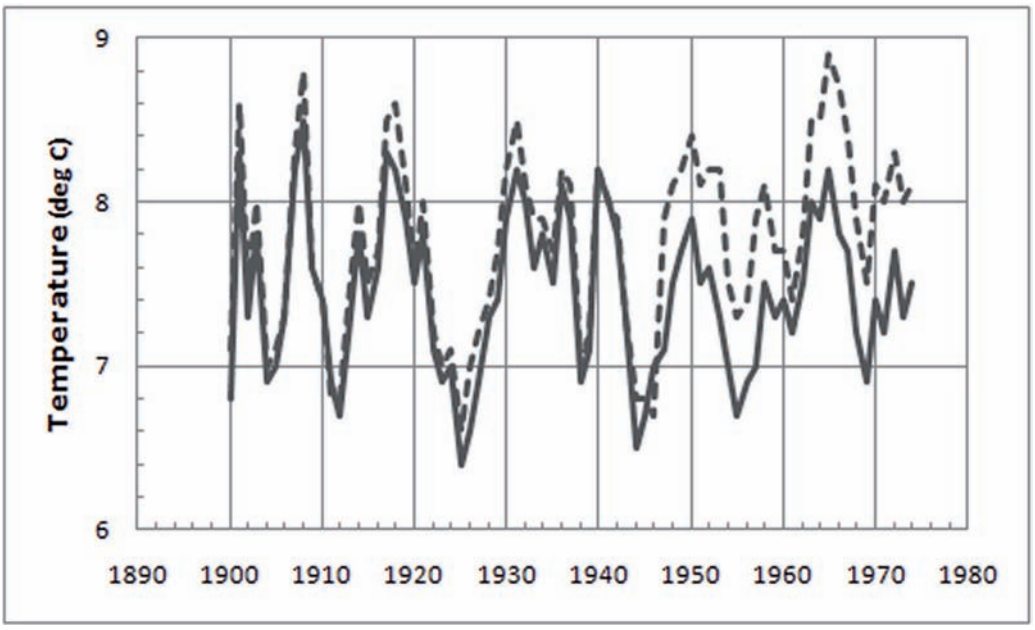


**Figure 5.1. Example of a candidate time series (dashed line) and a reference (solid line) time series**

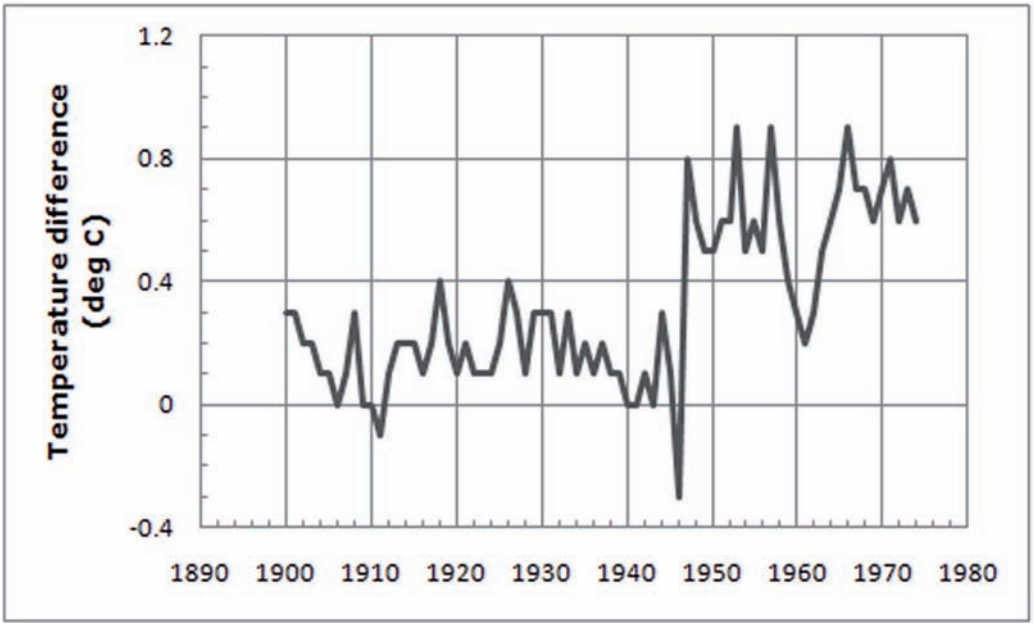


**Figure 5.2. Example of a difference time series**

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 a runs tests, 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 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)

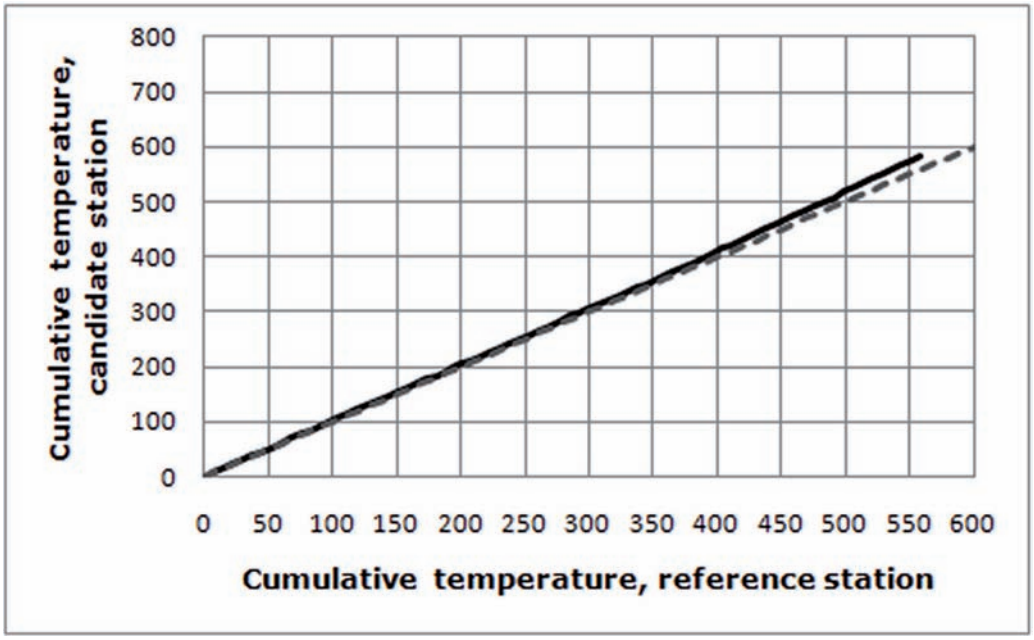


**Figure 5.3. Example of a double-mass graph with the dashed line representing a slope of 1**

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 often have a unique set of weather conditions creating them. With few extreme data points 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 those from phenology, observation journals, or ice freeze and thaw dates. When using such strategies, one also has 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. The adjusted series accounting for the time of observation bias, as one might predict, shows more warming over time than the unadjusted dataset.

More complete descriptions of several widely used tests are available in the *Guidelines on Climate Metadata and Homogenization* (WMO/TD-No. 1186) and in several of the references listed at the end of this chapter. 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. Some single values may remain incorrect or made even worse by the homogenization, however.

## 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 section 4.4.1 so that statistical methods can be exploited. A theoretical frequency distribution can be fit 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 are called degrees 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 taken. 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 tests 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.

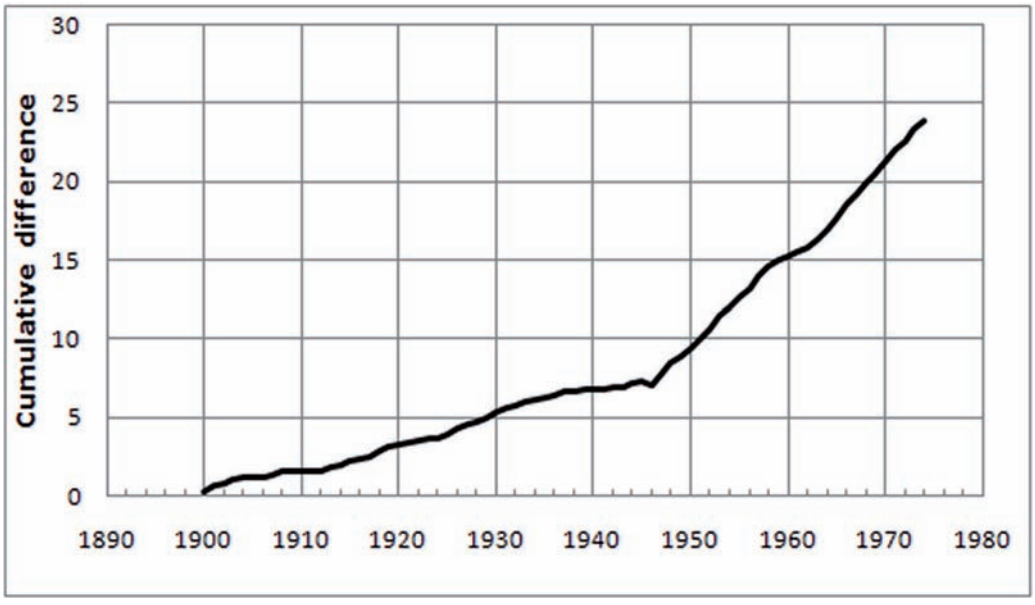


**Figure 5.4. Example of a residual double-mass graph**

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 section 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 transform 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 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 section 4.4) to inferential tests of normality.

Prior to applying any transformation, an analyst must make certain 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 utilized 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 compresses 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, and 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. 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, however. 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. The transformed distributions will likely become more complex to interpret physically, however, 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 section 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 patterns are found, then the model is adjusted to incorporate the patterns.

Time series in climatology have been analysed mainly with harmonic and spectral analysis techniques that decompose a series into time domain or frequency domain components. A critical assumption of these models is that of stationarity (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 section 5.2).

Gabor and wavelet analysis are extensions of the classical techniques of spectral analysis. By allowing subintervals of a time series to be modelled with different scales or resolutions, the condition of stationarity is relaxed. 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, they are 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 do have some limitations. The most important limitation for 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 improper model), and periodicity inherent in the data but not modelled. Split validation is effective in detecting model overfitting. Split validation refers to 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 or a compilation of observations of one element at different points in 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 and can be classified, if the elements fall into groups, and if interdependence exists among elements. Such datasets are also used to test hypotheses about the data. The time order of the observations is generally not a consideration; 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. In the meteorological and climatological literature, factor analysis is often called rotated principal components analysis. It is similar to principal components analysis except that the factors are not uncorrelated. Since a factor may represent observations from 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 section 6.7.3), in determining homogeneous regions for flood forecasting in an ungauged basin, and in reconstructing spatial wind patterns from pressure fields.

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 from 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 a 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 section 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 affects the degree

of smoothing. Estimation near the boundaries of the data is problematic as well. Outliers can severely affect a spline fit, especially in regions with few observations.

A range of more sophisticated, often non-parametric smoothers, are 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 data 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 be 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 also has 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 of 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 as 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 neighbours 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 to 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 to 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 assumptions are not fulfilled, particularly when the variances of the series at the two stations are not equal for the method using the differences, these techniques should not be used. More complex physical consistency 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 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, for example: 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 and 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 are 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.

Since meteorological or 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 section 6.7). As the computer power and storage capacity they require becomes more readily available, these models are being used more widely in climate monitoring, and especially to estimate the value of climate elements in areas remote from actual observations (see section 5.13 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 timescales 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 section 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.

### 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 be in error. Furthermore, the error could be serious and lead to incorrect conclusions. In climatological analysis, model assumptions often are 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 in error. 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 that this assumption will hold, and consequently the interpolated values could be in error.

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 spatial interpolation 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 to compare 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 mapped for visual inspection. Cross-validation offers quantitative insights into how any estimation method performs. An analysis of the spatial 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 streamflows 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 have assumptions such as stationarity and independence of data values, as discussed in Chapter 4. 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 fitted by an extreme value distribution model 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. The generalized

Pareto frequency distribution is usually suitable for fitting data beyond a threshold.

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 therefore it is 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 account for the decadal-scale changes in climate observed in the recent past, as well as for 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) is a publication that provides more insight into how one should account for 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 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, but 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 include this fact when disseminating the results to users. It may be also be 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 reanalyse 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 section 3.3).

Basic data management tools provide a wide variety of operations with which to make the data convenient for processing and analysis. These operations include sorting, adding data, subsetting 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 usually be saved, exported and transformed, and thus used for any further analysis and processing.

The graphical tools contained in statistical packages include the creation of two- and three-dimensional graphs, the capability to edit the graphs, and the capability to save the graphs in 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 the 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 these selections. This flexibility is useful, for example, 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 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.

## 5.14 REFERENCES AND ADDITIONAL READING

### 5.14.1 WMO publications

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

———, 1981: *Selection of Distribution Types for Extremes of Precipitation* (OHR-No. 15, WMO/TN-No. 560), Geneva.

———, 1986: *Manual for Estimation of Probable Maximum Precipitation* (OHR-No. 1, WMO/TN-No. 332), Geneva.

———, 1990: *Extremes and Design Values in Climatology* (WCAP-No. 14, WMO/TD-No. 386), 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.
———, 1997: *Progress Reports to CCl on Statistical Methods* (WCDMP-No. 32, WMO/TD-No. 834), Geneva.
———, 1999: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data* (Budapest, Hungary, 9–13 November 1998) (WCDMP-No. 41, WMO/TD-No. 962), Geneva .
———, 2002: *Guide to the GCOS Surface and Upper-Air Networks: GSN AND GUAN* (GCOS-No. 73, WMO/TD-No. 1106), Geneva.
———, 2003: *Guidelines on Climate Metadata and Homogenization* (WCDMP-No. 53, WMO/TD-No. 1186), Geneva.
———, 2008: *Deficiencies and* Constraints *in the DARE Mission over the Eastern Mediterranean: MEDARE Workshop Proceedings* (WCDMP-No. 67, WMO/TD-No. 1432), 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.

### 5.14.2 Additional reading

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Int. J. Climatol.*, 6:661–675.
Angel, J.R., W.R. Easterling and S.W. Kirtsch, 1993: Towards defining appropriate averaging periods for climate normals. *Climatol. Bull.*, 27:29–44.
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 Resource 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. *Int. J. Climatol.,* 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. *J. Appl. Meteorol.*, 33:140–158.
Della-Marta, P.M. and H. Wanner, 2006: Method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, 19:4179–4197.
Deutsch, C.V. and A.G. Journel, 1998: *GSLIB: Geostatistical Software Library and User's Guide*. Second edition. Oxford, Oxford University Press.
Dixon, K.W. and M.D. Shulman, 1984: A statistical evaluation of the predictive abilities of climatic averages. *J. Clim. Appl. Meteorol.*, 23:1542–1552.
Environmental Systems Research Institute (ESRI), 2008: *ArcGIS Geographical Information System*. Redlands, ESRI.
Fisher, R.A. and L.H.C. Tippet, 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 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 Res.*, 19:193–212.
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. *J. Geophys. Res.*, 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, IPCC.
International ad hoc Detection and Attribution Group (IDAG), 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate*, 18:1291–1314.
Jenkinson, A.F., 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.*, 81:158–171.
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.
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.
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. *Bull. Amer. Meteorol. Soc.*, 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. *J. Geophys. Res.*, 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. *J. Clim. Appl. Meteorol.*, 25:145–160.

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

Kohler, M.A., 1949: Double-mass analysis for testing the consistency of records for making adjustments. *Bull. Amer. Meteorol. Soc.*, 30:188–189.

Mann, M. E., 2004: On smoothing potentially non-stationary climate time series. *Geophys. Res. Lett.*, 31: L07214, DOI:10.1029/2004GL019569.

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 Climate Research Division, Environment Canada, and National Climatic Data Center, NOAA. Asheville, NCDC.

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. *Int. J. Climatol.*, 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, ICPO.

Palus, M. and D. Novotna, 2006. Quasi-biennual oscillations extracted from the monthly NAO index and temperature records are phase-synchronized. *Nonlin. Process. Geophys.*, 13:287–296.

Peterson, T. C. et al., 2002. Recent changes in climate extremes in the Caribbean Region. *J. Geophys. Res.*, 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. *J. Climate,* 20:4065–4095.

Sen, P.K., 1968: Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Stat. Assoc.*, 63:1379–1389.

Sensoy, S., T.C. Peterson, L.V. Alexander and X. Zhang, 2007: Enhancing Middle East climate change monitoring and indices. *Bull. Amer. Meteorol. Soc.*, 88:1249–1254.

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

Torrence, C. and G.P. Compo, 1998. A practical guide to wavelet analysis. *Bull. Amer. Meteorol. Soc.*, 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, A.C.M. Beljaars, 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. *Q. J. R. Meteorol. Soc.*, 131:2961–3012.

Vincent, L.A., X. Zhang, B.R. Bonsal, and W.D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. 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 Rev.*, 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. *J. 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. *J. Appl. Meteorol. Climatol.*, 47:2423–2444.

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

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. *J. Hydrol.*, 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. *J. Geophys. Res.*, 110: D22104, DOI:10.1029/2005JD006181.