Statistical Descriptors of Climate

Nathaniel B. Guttman
National Climate Data Center
Asheville, NC 28801

Abstract

An adequate description of climate is required to meet the informational needs of planners and policy-makers who use climate as a factor in their decision-making processes. Because normals have become firmly entrenched as a descriptor of climate, their history and their perception by the public are discussed. An "exploratory data analysis" approach is suggested.

1. Introduction

In his classic handbook, Hann (1903) described *climate* as the sum total of the meteorological phenomena that characterize the average condition of the atmosphere at any one place on the earth's surface. Although Hann recognized the importance of departures from average conditions, he emphasized that climate should be described by average weather conditions. During the 1800s and the first half of the 1900s, this view dominated the work in climatology. Smith (1975) states that during this time period, description and classification of climate, often based on means, prevailed. The concept of normals also evolved and matured during this time period.

A gradual change in thinking has taken place during the past few decades. Information about average climatic conditions is still important, but descriptions of climatic variability have become prominent. The importance of climatic variability lies in the impacts of climate on soil, water, food, energy, and shelter—the basic ingredients of survival. Planners and policymakers who use climate as a factor in their decision-making processes require an adequate description or assessment of climate.

The history of normals and perception of normals by the public are discussed first; the inadequacies of normals in meeting the climatic information requirements of the users are then outlined, and an "exploratory data analysis" approach to describing and modeling climate is presented. The results of the model should give the planner practical information about the underlying structure of the climate system of interest.

2. Background of normals

The term "normal" first appeared in the meteorological literature in 1840 in a monograph by Dove on temperature variations over the globe (Landsberg 1955). The word was used by Dove in several papers in three different contexts: (1) as a reference value obtained for a given latitude belt by averaging all observations in that belt, (2) as a reference locality with a long observational record for comparisons in time and space, and (3) as equivalent to the average or mean of a long series of observations.

The third context survived during the latter half of the 19th century. The International Meteorological Committee in 1872 resolved to compile mean values over a uniform period in order to assure comparability between data collected at various stations. According to Landsberg (1955, 1972, 1975), the doctrine gradually developed that climate is essentially constant during intervals that are long compared to human experience. It was assumed that long-term averages would converge to this stable value or normal. International agreements eventually led to the compromise that the appropriate interval for computing a normal would be 30 years. In 1935 it was also agreed that 1901 to 1930 would be the period for computing normals.

One of the problems of these normals is that many places in the world did not begin observations in 1901. Another problem is that the notion of a "stable climate" has, over the last few decades, become obsolete. The World Meteorological Organization, which eventually succeeded the International Meteorological Committee, has attempted to rectify these problems by using a sliding time scale to compute normals (Jagannathan et al. 1967). *Normals* are now defined as "period averages computed for a uniform and relatively long period comprising at least three consecutive 10-year periods" (WMO 1984). *Climatological standard normals* are "averages of climatological data computed for the following consecutive periods of 30 years: January 1, 1901 to December 31, 1930, January 1, 1931 to December 31, 1960, etc." (WMO 1984). Normals are computed every decade in an attempt to keep up with any climatic changes that may take place.

Climatologists generally understand that a normal is simply an average of a climatic element over thirty years, and that departures from normal are the difference between currently observed values and that particular 30-year average. It is also generally recognized by climatologists that a "normal" value is usually not the most frequent value (mode) nor the value above which half the cases fall (median). The general public, however, has a tendency to perceive the normal as ordinary or frequent (Landsberg 1955, 1972). The perception probably arises from the common usage of the word normal: "(1) conforming, adhering to, or constituting a usual or typical standard, level or type, (2) the usual or expected state, form, amount or degree" (DeVinne 1982).

The improper interpretation of normals by the public has a profound effect on the results of applying the values that are published each decade. Any large deviations from the temporal averages are thought to be indicative of climatic change, even though they may be only or part of the "noise" pattern (Landsberg 1975). As stated by McKay (1975), planners make decisions based on the assumption that a 30-year average has predictive value. Since the assumption is doubtful, McKay claims that those decisions must ultimately be in error and may lead to unfortunate consequences.

3. Normals and needs

Historically, normals have been used for two purposes—comparison and prediction. The former usage allows the analyst to evaluate and assess the deviation of a value of a climatic element from that for a reference period (Jagannathan et al. 1967; Saxton 1979). Normals also allow comparison in synoptic analysis, i.e., spatial comparisons at a given time. For these purposes, the period of record should be uniform for all stations. The normals as they have been previously defined and published meet the needs of those making these kinds of comparisons. It is emphasized, however, that these comparisons imply very little about climatic change, non-random fluctuations, or extremes. They are simply an assessment of deviations from a reference value.

The second use of normals is prediction from a 30-year average. *Prediction* is defined, for our purposes here, as planning or assessing risks based on the likelihood of recurring climatic scenarios. It is instructive to examine what is being predicted in several areas of activity so that commonalities can be drawn.

In 1983 a workshop was convened to discuss the interactions of climate and energy. Landsberg (1984) in his keynote paper stressed that industries are not only concerned with moderate weather and climate fluctuations, but also with extremes of the climatic elements. The oil industry (Leavitt 1984) needs precise estimates of the range of climatic variation and, in particular, extremes. This climatic information is used in planning for oil exploration and in estimating energy use. The gas industry requires a long-term weather history to plan future supply needs, facilities, and storage, and also requires means and extremes to determine price structures (Laurmann 1984). Climatic descriptions including averages, variability, extremes, and special events are needed by the coal industry for planning reclamation activities and avoiding weather-related shutdowns (Wilson 1984; McKee and Doesken 1984). The electric utilities require predictions of means and extremes for determining rate structures, baseline and peak loads, and facility siting (Rotty 1984).

The climatic information needs of agribusiness are summarized by Lamb et al. (1985). Historical data such as normals are used by the industry in assessing frost risks, estimating crop yields, analyzing supply and demand relationships, developing marketing strategies, planning planting schedules and herbicide applications, evaluating chemical-product trials, and scheduling of financial borrowing and investment activity. Most of the information is used for general background as opposed to making specific decisions. The needs, however, include predictions of climate, detailed information for input to crop yield models, and estimates of variability for design and planning of operations. The need for analytical measures of climatic variability were also recently mentioned by many authors, e.g., Lehman (1987), Todorov (1985), Smolander and Lappi (1985), Russell (1984), and Stern and Coe (1982). An extensive annotated bibliography on the use of climatic variables that are agriculturally sensitive is given by McQuigg (1975).

Other areas where climatic assessments are made are the building industry and architecture. Architectural design criteria include temperature means, extremes and variability; solar radiation variability; average and extreme wind (Simiu and Lozier 1975); and precipitation averages and extremes (Watson and Labs 1983; Marsh 1977; Olgyay 1963; Aronin 1953). These climatic factors enter into the placement of building openings such as windows, directional orientation, landscaping, building shape, etc. The materials used in construction as well as the mechanical systems used to heat and cool a building are determined partially by the risk of exceeding values of specified climatic elements (Ecodyne 1980; Lunde 1980; ASHRAE 1985). The need for climatic information is based on the goal of designing and building safe and comfortable structures with a predetermined level of risk of collapse or discomfort.

Applying climatic information to the prediction of future events in virtually all disciplines is similar to the examples just described. The common feature is the assessment of risk or prediction error. The predictive value of normals was extensively studied over two decades ago by Enger (1959), Court (1967, 1968a–c), and Slusser (1968). After his 3-year study, Court (1968d) stated, "Climatic normals. . .are extremely inefficient for the primary use to which they are put: estimating future conditions" and ". . .the concept of climatic normal should be abandoned in practical climatology." Normals provide one measure of "central tendency" for a specified period of record, but users require a complete statistical description of a climatic element, i.e., central tendencies, extremes, variability, and durations, in order to assess risk.

4. A satisfactory description of climate

The limited applicability of normals in meeting the needs of users who are interested in predicting future atmospheric events was outlined in the previous section. The following discussion defines what might constitute an adequate portrayal of climate and proposes an approach for obtaining the description.

According to Jagannathan et al. (1967), a satisfactory specification of the climate of a region or locality includes statistical measures such as the mean, range, frequency, and variations of several climatic elements, as well as the sequential character of the elements. The usually important climatic elements are wind, temperature, moisture, and pressure; phenomena observed within the atmosphere, such as rainfall, snow, and cloud amount; and sunshine and radiation balance. Jagannathan et al. (1967) further state that "climatology should provide information, as accurately as the available observations warrant, about the state and behavior of the atmosphere." The aim is not only to compute normals, but also to abstract, from the past observational record, the underlying patterns characterizing the atmospheric environment. These patterns become the basis for risk assessments and prognostications discussed in the previous section.

The collection of historical weather data, i.e., the observational record, is an incomplete log of complex temporal and spatial interactions in the global environment. Obviously, data do not exist for every geographical point for all time. In addition, the data are subject to errors of instrumentation, observing, and processing. The chronological series of data can therefore be considered as a realization or sample of the underlying physical processes that control climate.

A basic approach in analyzing the record of observations for the purpose of describing climate includes three steps and is known as exploratory data analysis (Tukey 1977). Andrews (1978) defines the approach as the manipulation, summarization, and display of data to uncover the underlying structure in the data. Judging from the literature, the procedures have only recently been formally applied to atmospheric data (Snijders 1986; Flueck et al. 1986; Zeger 1985), although consulting meteorologists and climatologists have probably used the analytical techniques informally for decades.

The first step of the analysis is examination. It is designed to uncover systematic errors, to exhibit overall patterns, and to show departures from an a priori contemplated structure. Graphical tools are appropriate for this phase. "Stem-and-leaf" displays, histograms, weather diagrams, quantile plots, cumulative-frequency distributions, maps, and time-series plots are some of the graphical tools that allow an evaluation of the structure of the data. A good knowledge of the data acquisition and measurement processes as well as the physics underlying the climatic element is invaluable in interpreting the graphics.

It is appropriate during this step to perform nonparametric tests to determine randomness, symmetry, trends, correlations, and differences among samples. Several of these tests are discussed by Siegel (1956). Descriptive statistics—those that compactly express the salient features of the observations—should also be computed. These statistics include the sample average, median, and variance, i.e., measures of "location" and dispersion.

The data examination gives the analyst information about inhomogeneities within the observational record. At this point the data should be adjusted to remove known systematic errors such as instrument and observer biases and gradual changes in site exposures (Brooks and Carruthers 1953); inhomogeneity caused by discontinuities in the observations, such as instrument changes and recalibrations, location changes of the observing site, and changes in observing methods; changes in computational procedures; environmental changes at the site such as urbanization; and data processing errors in calculating, coding, transcribing, and transmitting. (An excellent review of errors in meteorological data is given by Filippov 1968.) Adjustment methods that compensate for some of these problems are described by, among others, Brooks and Carruthers (1953), Landsberg (1958), Conrad and Pollak (1950), Thom (1966), and Craddock (1981). After the known errors have been removed from the data, the corrected observational record should once again be examined and descriptive statistics should be recomputed.

Of importance is that the examination step makes no assumptions about statistical models. It provides information that gives direction to "cleaning up" the data as well as allowing the analyst to postulate reasonable models.

The second step, estimation, relies on the formal procedure of inferential statistical analysis and is concerned with the process of using data to make decisions about general situations on the basis of incomplete information (Mood and Graybill 1963). The analysis involves selecting a statistical model, checking the reasonableness of the model, and then drawing conclusions from the model (Hoel 1962). The model chosen assumes the existence of a signal such as a trend, oscillation, persistence, dependence, or frequency distribution in the observations. Specification of the model is based on the descriptive statistics and patterns that are identified and interpreted during the examination step.

Evaluation of the model is based on its "goodness-of-fit" to the real-world pattern being described, by the fulfillment of assumptions upon which the model is based, and by the degree of confidence that can be placed on the statistical inferences drawn from the model. If the model incorrectly describes a feature inherent in the data, then another model should be chosen, applied, and evaluated.

A statistical model that is deemed reasonable merely describes what is a feature within data. It does not explain why the feature occurs. The "what" becomes the raw material for developing physical models that explain "why." Statistical models describe patterns in a given data set. Physical models, on the other hand, explain the underlying physics of observed patterns and are not dependent upon any one data set.

Once a statistical model is determined to reasonably describe a data feature, values calculated from the model are removed from the data. The removal process, which is the third step in the exploratory data analysis, typically involves a subtraction or other transformation based on the statistical model. With the first gross component of the structure removed, the reduced or residual data may be examined for finer structure. This finer structure may again be modeled, estimated, and removed. The exploratory data analysis process may be repeated until all of the apparent structure has been removed and the residual data appear to be random, or patternless.

An example based on Guttman and Plantico (1987) illustrates the exploratory data analysis concept. Examination of 1951–1980 average daily maximum temperatures at stations in the eastern United States showed an annual cycle in the data. A smooth curve through average monthly maximum temperatures at each station was chosen as a model to represent the observed temperature patterns. Evaluation of the model led to the conclusion that the smooth curve is reasonable.

For each station, daily values of the smooth curve were subtracted from the observed data, i.e., the modeled values were removed from the data. The residual data were examined and found to exhibit a one-day persistence. A first-order, autoregressive model was chosen to describe the persistence in the data. Evaluation of this model showed that the first-order autoregression did describe the persistence. Note that the autoregression describes, but does not explain, the persistence; "what," but not "why," is determined.

The values obtained using the autoregressive model were subtracted from the residual data yielding a second, residual data set. Examination revealed nonrandomness and apparent quasi-periodic fluctuations in this second set. These fluctuations are sometimes referred to as singularities, i.e., certain times of the year when temperatures are considered anomalous. (An example is the January thaw.) A spectral-analysis model was chosen to describe the nonrandomness; evaluation showed the model was inappropriate. The search for a reasonable description of the second, residual data set is continuing.

The observed temperature data have been partially described by the sum of two components, the annual cycle and the autoregression. A "complete" description requires additional iterations of the exploratory data analysis approach. The advantage of the approach is the isolation of independent components upon which statistical inferences can be made and upon which research efforts to develop physical models can be focused.

A complete statistical model of an observational record is composed of all the components used to describe the systematic portion of the record, as well as the statistical description of the residual portion of the record. If based on sound statistical principles, the model represents, but does not explain, the underlying statistical processes that the data measure. Within a probabilistic framework, the model gives information appropriate for estimating the likelihood of occurrence of future events and for estimating risk.

5. Supplementing normals

A complete exploratory data analysis is a difficult, time consuming, expensive process. Results from a partial analysis will, however, provide the user of climatic data with more information than is currently readily available. Current routinely published, summarized information includes averages, extremes, normals, and departures from normals. Limited histories of station location and instrumentation are also available.

The examination step of an exploratory data analysis yields descriptions of climatic data that supplement normals. It identifies distributional characteristics of climatic data such as symmetry, independence of observations, and variability. Publication of descriptors such as standard deviations for symmetric distributions, medians and quantiles for asymmetric distributions, and correlation coefficients for dependent data would provide users with more information than is conveyed by publishing only one measure of central tendency, i.e., an average.

Station histories could be improved from the knowledge gained from the examination step. Detailed historical documentation of non-climatic effects such as computation procedures, instrument response, site exposure, changes in recording times of observations, and data-correction procedures would help an analyst assess the homogeneity of a climatological data series.

Publication of descriptive statistics, observing-system characteristics, and processing procedures will not solve the user's problem of predicting future events from the climatological record. It will, however, lead to a better understanding of the nature of climatic parameters. This understanding becomes the basis for developing a set of characteristics from which predictions can be made.

6. Summary

Normals have become firmly entrenched as a descriptor of climate. They were developed for comparative purposes. For predictive purposes, a more complete description of climate is required than is provided by normals. An exploratory data analysis approach is suggested to provide the climatic descriptions necessary for developing predictive relationships. The approach is based on the modeling of systematic phenomena and of probabilistic estimates of random, patternless phenomena. It necessitates validation and clean-up of data, utilizes the information available in the observational record, and results in descriptions for specific user applications.

References

Andrews, D. F. 1978. Data analysis, exploratory. *International Encyclopedia of Statistics*, ed. W. H. Kruskal and J. M. Tanner. New York: Free Press.

Aronin, J. E. 1953. *Climate and Architecture*. New York: Reinhold.

ASHRAE. 1985. *ASHRAE Handbook, 1985 Fundamentals*. Atlanta: Amer. Soc. Heating, Refrigerating and Air-Conditioning Engineers.

Brooks, C. E. P., and N. Carruthers. 1953. *Handbook of Statistical Methods in Meteorology*. London: Her Majesty's Stationery Office.

Conrad, V., and L. W. Pollack. 1950. *Methods in Climatology*. Cambridge: Harvard Univ. Press.

Court, A. 1967. Climatic normals as predictors, Part 1: Background. AFCRL 67-0313. Air Force Cambridge Research Laboratories, Bedford, MA.

——— 1968a. Climatic normals as predictors, Part 2: Extension. AFCRL 68-0673. Air Force Cambridge Research Laboratories, Bedford, MA.

——— 1968b. Climatic normals as predictors, Part 4: Verification. AFCRL 69-0002. Air Force Cambridge Research Laboratories, Bedford, MA.

——— 1968c. Climatic normals as predictors, Part 5: Conclusion. AFCRL 69-0003. Air Force Cambridge Research Laboratories, Bedford, MA.

——— 1968d. Climatic normals are inefficient. AMS Conference and Workshop on Applied Climatology, October 29–31, 1968, Asheville, NC.

Craddock, J. M. 1981. Monitoring a changing climate series. *J. Clim.* **1:** 333–343.

DeVinne, P. B., ed. 1982. *The American Heritage Dictionary, 2nd Coll. Ed.* Boston: Houghton Mifflin.

Ecodyne. 1980. *Weather Data Handbook*. New York: McGraw-Hill.

Enger, I. 1959. Optimum length of record for climatological estimates of temperature. *J. Geophys. Res.* **64:** 779–787.

Filippov, V. V. 1968. Quality control procedures for meteorological data. World Weather Watch Planning Report No. 26. World Meteorological Organization, Geneva.

Flueck, J. A., W. L. Woodley, A. G. Barston, and T. J. Brown. 1986. A further assessment of treatment effects in the Florida Area Cumulus Experiment through guided linear modeling. *J. Clim. Appl. Meteor.* **25:** 546–564.

Guttman, N. B., and M. S. Plantico. 1987. Climatic temperature normals. *J. Clim. Appl. Meteor.* **26:** 1428–1435.

Hann, J. 1903. *Handbook of Climatology, Part I*. R. DeC. Ward, trans. London: Macmillan.

Hoel, P. G. 1962. *Introduction to Mathematical Statistics*. New York: Wiley and Sons.

Jagannathan, P., R. Arley, H. ten Kate, and M. V. Savarina. 1967. A note on climatological normals. Tech. Note No. 84, WMO-No. 208.TP.108. World Meteorological Society, Geneva, Switzerland.

Lamb, P. J., S. T. Sonka, and S. A. Changnon. 1985. Use of climate information by U.S. agribusiness. NOAA Tech. Rpt. (NCPO-001). U.S. Department of Commerce, NOAA, Washington, D.C.

Landsberg, H. E. 1955. Weather "normals" and normal weather. *Weekly Wea. Crop Bull.* **42:** 7–8.

Landsberg, H. E. 1958. *Physical Climatology* (*2nd ed.*). DuBois, PA: Gray Printing.

——— 1972. Weather "normals" and normal weather. *Environmental Data Service*. October 1972: 8–13. National Oceanic and Atmospheric Administration.

——— 1975. The definition and determination of climatic changes, fluctuations and outlooks. *2nd Carolina Geographical Symposium, Atmospheric Quality and Climatic Change*. Ed. R. J. Kopec. Chapel Hill: U. North Carolina.

——— 1984. Overview: Climate and energy interactions. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

Laurmann, J. A. 1984. Summary report of the panel on climate and the gas industry. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

Leavitt, P. R. 1984. Improving the utility of climatic data to the oil industry. Proc. of the DOE/Industry Workshop on the Interaction of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

Lehman, R. L. 1987. Probability distributions of monthly degree day variables at U.S. Stations, Part I: Estimating the mean value and variance from temperature data. *J. Clim. Appl. Meteor.* **26:** 329–340.

Lunde, P. 1980. *Solar Thermal Engineering.* New York: Wiley and Sons.

Marsh, P. 1977. *Air and Rain Penetration of Buildings.* Lancaster, England: Construction Press, Ltd.

McKay, G. A. 1975. Future climate and decision-making. *2nd Carolina Geographical Symposium, Atmospheric Quality and Climate Change.* Ed. R. J. Kopec. Chapel Hill: U. of North Carolina.

McKee, T. B., and N. J. Doesken. 1984. An assessment of climate information to serve the energy industry in Colorado. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

McQuigg, J. D. 1975. *Economic Impacts of Weather Variability.* Department of Atmospheric Science, University of Missouri-Columbia.

Mood, A. M., and F. A. Graybill. 1963. *Introduction to the Theory of Statistics.* New York: McGraw-Hill.

Olgyay, V. 1963. *Design with Climate.* Princeton: Princeton University Press.

Rotty, R. M. 1984. Summary report of the panel on climate and the electric utilities. Proc. of the DOE/Industry Workshop on the Interactions of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

Russell, J. S. 1984. Mathematical representation of related period variables using Fourier series and projections of vector functions: Application to the biological sciences. *Agric. For. Meteor.* **33:** 215–224.

Saxton, D. W. 1979. Climate normals in the Greenland Sea area. *J. Appl. Meteor.* **18:** 1222–1223.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

Simiu, E., and D. W. Lozier. 1975. The buffeting of tall structures by strong winds. NBS Building Ser. No. 74. U.S. Department of Commerce, National Bureau of Standards, Washington, D.C.

Slusser, W. F. 1968. Climatic normals as predictors, Part 3: Median vs. Mean. AFCRL 68-0255. Air Force Cambridge Research Laboratories, Bedford, MA.

Smith, K. 1975. *Principles of Applied Climatology.* New York: Wiley.

Smolander, H., and J. Lappi. 1985. Integration of a nonlinear function in a changing environment: Estimating photosynthesis using mean and variance of radiation. *Agric. For. Meteor.* **34:** 83–91.

Snijders, T. A. 1986. Interstation correlations and nonstationarity of Burkina Faso rainfall. *J. Clim. Appl. Meteor.* **25:** 524–531.

Stern, R. D., and R. Coe. 1982. The use of rainfall models in agricultural planning. *Agric. For. Meteor.* **26:** 35–50.

Thom, H. C. S. 1966. Some methods of climatological analysis. WMO No. 199. TP. 103. World Meteorological Organization, Geneva.

Todorov, A. V. 1985. Sahel: The changing rainfall regime and the "normals" used for its assessment. *J. Clim. Appl. Meteor.* **24:** 97–107.

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

Watson, D., and K. Labs. 1983. *Climatic Design.* New York: McGraw-Hill.

Wilson, J. S. 1984. Summary report of the panel on climate and the coal industry. Proc. of the DOE/Industry Workshop on the Interaction of Climate and Energy. CONF-830669. U.S. Department of Energy, Washington, D.C.

World Meteorological Organization. 1984. Technical regulations, Vol. I. WMO-NO. 49. Geneva, Switzerland.

Zeger, S. L. 1985. Exploring an ozone spatial time series in the frequency domain. *J. Am. Stat. Assn.* **80:** 323–331. •