Progress toward the better estimation of regional evapotranspiration

David Winter Meek
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/rtd

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Atmospheric Sciences Commons, and the Hydrology Commons

Recommended Citation
Meek, David Winter, "Progress toward the better estimation of regional evapotranspiration " (1997). Retrospective Theses and Dissertations. 11728.
http://lib.dr.iastate.edu/rtd/11728
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700     800/521-0600
Progress toward the better estimation of regional evapotranspiration

by

David Winter Meek

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Agricultural Meteorology; Water Resources

Major Professor: Jerry L. Hatfield

Iowa State University

Ames, Iowa

1997
This is to certify that the Doctoral dissertation of

David Winter Meek

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major/Professor

Signature was redacted for privacy.

For the Co-major Program

Signature was redacted for privacy.

For the Co-major Program

Signature was redacted for privacy.

For the Graduate College
iv

APPENDIX B: SEMIVARIOGRAMS AND CORRELOGRAMS FOR ALL 1990 ALBEDO DATA SETS 137

GENERAL BIBLIOGRAPHY 169

ACKNOWLEDGMENTS 172

VITA 173
1

GENERAL INTRODUCTION

Basic methodology for improving the regional estimation of net radiation and ultimately for the regional estimation of evaporation and evapotranspiration is needed to assess variation over many different spatial and temporal scales. This introduction provides relevant scientific background on a descriptive level, rationale for the dissertation, and an outline of the dissertation explaining the format, organization, and content.

1. Scientific background

Evaporation is the conversion of liquid water (or sublimation for ice) into water vapor. Evaporation involves a cooling process because the fastest (most energetic) molecules escape the liquid to become vapor. Hence, the average kinetic energy (i.e., temperature) of the remaining molecules is lowered. Transpiration is the vaporization of water through the stomata in plants. Evapotranspiration combines the evaporation from the soil surface and small water surfaces along with transpiration by vegetation. The reason the terms are combined is to simplify the estimate and because both processes are occurring simultaneously in vegetated surfaces. The terms evaporation and evapotranspiration are often used interchangeably but evaporation can be considered generic for all vaporization processes unless more restrictive specifications are stipulated (Brutsaert 1982).

For small-area studies a water balance or budget is often employed to estimate evapotranspiration. In this scheme, evapotranspiration is a portion of the rainfall returned to the atmosphere by vaporization processes. On a regional scale, a water-budget approach is impractical because necessary measurements are generally labor-intensive and probably site specific. A related concept, surface energy balance or budget is more suitable for regional estimation of evapotranspiration. It is an expression of energy conservation (i.e., the first law of thermodynamics).
The primary source of energy for the hydrologic cycle, as with most other natural terrestrial processes, is the global solar radiation that is not reflected away from the surface (a.k.a. "net shortwave radiation"). The land and atmosphere radiate longwave energy according to their absolute temperatures. Throughout the year at midlatitudes outgoing longwave radiation at the land surface is generally somewhat in excess of incoming on a daily basis. When the situation reverses incoming longwave radiation from the atmosphere can also be an energy source.

At the land surface radiative transfer processes occur that result in energy exchanges such as evapotranspiration and heating the environment. At all times a balance forms between all the shortwave and longwave radiation components. This net radiant energy is generally the only significant energy source at the land surface in extensively cropped regions. It is available for partitioning into various uses (energy sinks) and is called "net radiation". The portion of this energy that is used for evapotranspiration is known as "latent heat". The energy used for heating the environment is called "sensible heat". The energy that passes through the surface to be stored in the ground is called "soil heat flux". There are other energy balance sources and sinks which are generally small in cropped areas and can usually be neglected.

When water is available evapotranspiration utilizes much of the available energy at the land surface. As previously stated, this available energy is generally only net radiation. Hence, determining surface net radiation is paramount in the estimation of evapotranspiration from an energy-balance approach over all scales of interest.

Two surface properties that play a role in the determination of net radiation are surface albedo and emissivity. Surface albedo is generally defined as the ratio of reflected to incoming global solar shortwave radiation. Emissivity is the ratio of power radiated per unit area of the land surface (generally with known material properties) to the power radiated from a black body at the same temperature; hence it effects the
outgoing longwave radiation from the land surface and resulting surface energy balance. Both surface properties are considered to be aerially extensive and large-scale data are often obtained through remote-sensing methodologies for the purpose of regionalizing land surface processes.

Two approaches are used to estimate areal or regional evapotranspiration. One approach scales up with process models using point-surface observations. The other scales down from a large area with generally longer duration and more empirically based climatic-process models. Hybrid methods also exist. No universally accepted or practiced methodology has yet emerged but both approaches seek to incorporate remote-sensing data.

One problem common to both approaches is that climate models operate at larger spatial scales and over longer temporal periods with generally longer time increments than do surface-hydrological-process models. Moreover relevant data, especially data for surface properties and processes for a given climate model cell, may be sparse and incomplete. Also, while the incoming radiation components can often be considered uniform over large areas, surface properties and processes can be extremely heterogeneous in time and space, especially within an area the size of a global climate cell. Another problem is that many process models or their components are based to some degree on site and time specific calibrations that may not be universally valid.

Some practical arguments favor following the scaling-up approach. Presently for evapotranspiration, knowledge and methodology are really satisfactory only for point scales and, perhaps, very small, well-studied land surface areas. All major data sets, especially those collected from planned, large-scale field hydrology experiments, include temporally dense point-source surface meteorological data at multiple sites, often on a grid. In addition, there often are atmospheric profile data and airborne and satellite remote-sensing data available, although both may be comparatively more temporally sparse.
2. Rationale for the dissertation

Understanding the surface processes involved in global and regional climate change, desertification, and flood and drought forecasting is of practical consequence to the hydrological sciences and their application to water resource management. One of the many important reasons for modeling the global climate is to assess the magnitude and distribution of possible global warming changes in many factors including precipitation and temperature. Such changes could be of considerable consequence in all facets of life but particularly to local and regional agriculture. For example, a sufficiently warmer and drier climate in the United States Corn Belt could be the ruin of the current rainfed farming practices and crop selection.

Applications to water resource management are also varied. In agriculture, particularly the arid and semiarid regions like those in the Western United States, irrigation scheduling and drainage management are vital. This management relies on modeling the water balance over large areas.

The need to estimate evapotranspiration over large areas is a common interest that links both scientific research and water resources management. In the United States Corn Belt example, normally evapotranspiration is less than rainfall. In such rainfed agricultural regions evapotranspiration is the major land-surface process that generally consumes most of the available water in the soil profile. As for water resource management in the United States, related research and public information expenditures have been and continue to be reduced and closely scrutinized (e.g., see Decker, 1994). As a result, over the past two decades, state, local, and private organizations have established networks of automated agricultural weather stations in many parts of the United States. One of their major functions is to provide local, presumably aerially valid, estimates of evapotranspiration.
Estimation of evapotranspiration begins with empirical procedures driven by the available data. Meteorological stations provide data that can be used to model what are known as "point-source" estimates of evapotranspiration which are generally considered valid only for a very small area depending on site and ambient conditions. One of the major problems with weather stations is the quality of the data collected.

Not surprisingly, concerns about the quality of surface meteorological data arose in a recent, locally conducted, watershed study. Automated surface meteorological stations that recorded real-time data (hourly and daily totals or averages) for several parameters were installed in two watershed research projects known as the Management Systems Evaluation Areas (MSEA) and were run by NSTL staff (USDA, 1994). These stations are similar to those in many of the previously mentioned networks. The need to devise and evaluate an automated data processing scheme for checking (being sure the value of a datum is reasonable) each datum for each one of the surface meteorological parameter data records was evident. The system was not to involve interactive statistics or utilize an expert system. Furthermore it should be useable by nonmeteorologists. This work resulted in paper 1. Partially as a result of working with the point source data, it became clear that there are basic problems in the estimation of regional evapotranspiration.

3. Dissertation organization

This dissertation consists of four papers that challenge extant practices, underlying assumptions and methodologies by offering methods, as examples of standards, that can be used in a scaling-up process for modeling net radiation and ultimately regional evapotranspiration.

The validity of point-source data is often taken for granted. When such data are checked the procedures used can be arbitrary and varied. The first paper "Data quality checking for single station meteorological databases" has been published in Agricultural and Forest Meteorology and is in the format for that journal. In this paper automated data-processing
rules based on climatology, physical principles, or instrument specifications were developed and evaluated for checking each variable in point-source surface meteorological data.

The second paper "A note on using the SAMSON database for estimating Linke turbidity factor" is in the format for the publications of the American Meteorological Society. It was included as a means of assessing turbidity data that were to be used in the third paper.

When needed point-source data are missing, not measured, or not directly measurable they are modeled, often with empirical or ad-hoc routines. The third paper "Estimation of maximum possible daily global shortwave solar radiation" is in the format for Agricultural and Forest Meteorology. In this paper broadband-radiative-transfer models were adapted, based on climatic low annual trends in the turbidity parameters, to estimate maximum possible daily global solar radiation at the land surface at an arbitrary location. This estimate, which cannot be directly measured, has multiple relevant uses. It can be used to screen daily global solar radiation measurements or to estimate global solar radiation or net radiation in the absence of these measurements. The model developed has a sounder physical basis than available alternatives.

In scaling up point-source surface measurements much attention has been paid to aggregation methods assuming the region or subregion of aggregation is known. The fourth paper "Scales of fluctuation for Iowa's albedo in 1990" is being prepared for Water Resources Research and follows the format for the American Geophysical Union's publications. In this paper spatial statistics were used to determine an averaging area for surface albedo data covering the state of Iowa for each of 16 periods of record throughout the 1990 growing season. The results were then assessed in the context of estimating regional energy balance.

A General Introduction and a Literature Review precede the papers while a General Conclusion follows them. The bibliography for the General Introduction and Literature Review follows the General Conclusion along
with two appendices. Appendix A includes graphs of clear day solar radiation data and all model curves for the remaining four sites reported in but not exhibited in the third paper. Appendix B comprises all semivariograms and correlograms from both the raw data sets and detrended data sets used in the fourth paper.
LITERATURE REVIEW

1. Research trends in regional evapotranspiration

Point and local scale evapotranspiration estimation has been the primary product of hydrological research and is often used for resource management. Areal or regional to global evapotranspiration estimation over a range of time scales is, however, needed to fully understand the impacts of land management on global climate change, various mesoscale climate models, and regional and national hydrological evaluations (Sellers et al. 1996; Henderson-Sellers et al. 1995; Henderson-Sellers et al. 1993; Bolle 1993b; Black et al. 1989; Eagleson 1986; or Hatfield 1985). There are difficulties with the estimation of areal or regional evapotranspiration so finding a means of directly measuring the quantity would be desirable but according to Hatfield (1990), direct, accurate, intensive, measurement is not feasible. Although presently many researchers believe that there is no universally applicable process model or operational method for estimating regional or areal evapotranspiration, research work on the problem is current and extensive (e.g., see Kirby 1996, Sellers et al. 1996, Henderson-Sellers et al. 1995; Bolle 1993b; Wilson 1989; or Jensen 1985).

In principle there are two conceptual approaches for modeling areal or regional evapotranspiration. They are scaling up, or up-scaling, and scaling down, or down-scaling. Both include assessing the spatial and temporal variability of the surface and its contribution to the energy balance. General definitions for each vary but generally scaling up involves finding suitable methods for extrapolating and/or aggregating point-source data over a defined domain (e.g., see Blöschl and Sivapalan 1995). This method is more widely utilized. Scaling down is generally the opposite process. Many, like Shuttleworth (1988), Bolle (1993), or Short et al. (1993), believe the need to reformulate processes at appropriate scales is primary to regional modeling. Much current research effort is
based on either one or a combination of both of these approaches (Bolle 1993).

One approach to scaling down is to theoretically start with the frame of reference at a regional scale and consider the meaning of observations or estimates from points within the region. Two prevailing empiricisms that have been tested in several environments are the Priestley and Taylor equation and the complementary relationship espoused by Bouchet and Morton (e.g., see Brutsaert 1982 or McNaughton and Spriggs 1989). Neither is completely satisfactory. Recent studies like that of McNaughton and Spriggs (1989) and older studies like those of Rouse et al. (1977) or Davies and Allen (1973) have shown reasonable support for the Priestley and Taylor equation but with the Priestley and Taylor coefficient showing variation with environmental conditions. McNaughton and Spriggs (1989), among others, have found that the complementary relationship is faulty; the premise that large-scale advection is independent of the surface energy balance is not true. Another more recent approach is to disaggregate one-dimensional atmospheric boundary layer estimates at the surface layer using available surface information (Guerra et al. 1993). More importantly, many researchers suggest that a satisfactory treatment is not yet manifest (e.g., see de Bruin 1989 or Sivapalan and Kalma 1995).

One approach to scaling up is to start with point estimates from weather stations and try to find a basis for the process. Point estimates are still the most accurate and best understood (e.g., see Jackson 1985 or Black and Spittlehouse 1989). Recent studies like MONSOON '90 (Kustas and Goodrich 1994) and FIFE (Sellers et al. 1988) have included a scaling-up approach that utilizes the surface observations, atmospheric soundings, and remote-sensing data. Scaling-up approaches can range between two methods and a variety of evapotranspiration and energy-budget components. One method involves interpolating energy balance inputs throughout a rasterized representation of a region, estimating evapotranspiration at each grid cell with a selected model, and then summing up the individual estimates over
all cells (e.g., see Hashmi et al. 1995). Such grid cells need not be of equal area but they need to be weighted by area (e.g., see Raupach, 1993). In addition, one can assume that a point estimate is valid for a region (e.g., see Hofstee et al. 1993). The other method involves estimating aggregated total or average values of the inputs over the region and then estimating regional evapotranspiration with a single calculation of a selected model as a function of the aggregated inputs (e.g., see Braden 1995 or Lhomme 1992). Kirby (1966), Henderson-Sellers et al. (1995), Bolle (1993), Becker (1989), and Hatfield (1985 and 1988) are among those who discuss the numerous problems, and often interconnected assumptions, inherent in utilizing available data with any of the approaches. Among the many extant problems in this subject, three seemed to be both fundamental, tractable, and related to the estimation of net radiation. One primary problem is the need to have reliable point-source data with the checking based on standardized procedures. Another one is the need for modeling parameters that are missing, not measured, or not measurable. For example in the modeling of daily net radiation, routines are needed that do not utilize arbitrary or locally empirical estimates for maximum possible daily solar radiation. The final one is the need to objectively determine averaging areas for the process and properties involved.

2. Quality assurance of data in meteorological databases

In many current projects in the climate-modeling sector, like those reported in Henderson-Sellers et al. (1993 and 1995) and Sellers et al. (1996), real surface data from many experiments were used to confirm their models; the reports mention data quality control protocols but the details are not reported. Furthermore the implementation of these protocols is fairly recent and limited to select data sets. In the operational agricultural sector, Meyer and Hubbard (1992), among others, review nonfederal government organizations that are establishing and operating automated weather station networks for multiple uses including
evapotranspiration estimation. In part the reason for this development is the diminishing role of the U.S. Government in providing data for agricultural and forest meteorology purposes (e.g., see Decker 1994). Hatfield and Fuchs (1990) maintain that viable evapotranspiration estimation for any given model with any given goal, is dependent on the accuracy of the basic surface meteorological observations at a given site. Hence the need to eliminate bad data is paramount.

Establishing a reasonable, minimum-level, practical quality control/quality assurance (QA/QC) program for the operation and maintenance of automated weather stations along with procedures for the collection and processing of the data can surely help improve the resulting measurement-based databases. Howell et al. (1984) outline the setup and routine operations at an individual station. Elwell et al. (1993) reviewed individual instrument problems, failure rates, and network operations and costs over a 10 year period. They found that, in general, less than one percent of the potentially available data were lost, radiometers deteriorated 5 to 7% over the first five years but maintained roughly constant performance thereafter, while relative humidity sensors deteriorated significantly before replacement resulting in portions of the data being useful only as a general guide to humidity conditions.

Checking the data in the data processing and database storage is also important. To some extent this task is being done, but by arbitrary methods. A basic automated standard for data from automated individual stations does not exist. Ashcroft et al. (1990a,b) used interactive graphics and human judgement for individual parameters. Hubbard (1988 and personal communication 1991) uses automated fixed range limits and regional maps for individual variables. Climatic-based or physically-based dynamic range limits and rate of change rules, like those proposed for stream gage hydrology by O’Brien and Keefer (1985), have not been examined for the purpose of checking surface meteorological data. While interactive graphical display is possible with such rules, they can be completely
automated with optional hard copy lists and graphs available for examination after the data are processed.

In Howell et al. (1984) daily net radiation was both measured and modeled. The model was similar to one in Jensen et al. (1971). Such a model uses most meteorological variables measured at a station. A way to realize a heuristic check of the data was to find agreement between the net radiation model and measurement. Unfortunately, net radiometers are a high maintenance instrument and they are not as widely distributed as global solar radiometers (Monteith and Unsworth 1990). Moreover, as in Dong et al. (1992), the emerging trend is to measure global solar radiation and model net radiation, even on an hourly basis, with reported comparisons to hourly measurements of ± 31 W m\(^{-2}\) for all hours and ± 23 W m\(^{-2}\) for hour angles greater than 10°. The hourly model of Dong et al. (1992) is more physically based than most others and can require the same kind of input variables as the previously mentioned daily model. Physically-based models for net radiation can require a quantity that is not directly measured, maximum possible global solar radiation.

3. Estimation of maximum possible daily global solar radiation

A related important need is to model basic parameters and/or properties not available in the database. Ideally this modeling of parameters should be done on a mechanistic level. Such a model should be applicable at any location and time within a region by just changing the input coordinates. Several empirical or simple broadband-radiative-transfer models have been proposed for estimating maximum possible daily global solar radiation (a.k.a "a clear-day curve"). Heermann et al. (1985) presented regression-based empirical sinusoidal and Gaussian annual-trend curves that could be used at any location within the continental United States. Their scaling of the maximum amplitude was arbitrary and the daily global solar radiation data selected did not come from the maximum daily value over a 30 year span. The WMO (1967) recommends a 30 year record for
determining normals and variation. Tracy et al. (1983) compared three, not completely physically comprehensive, broadband-radiative-transfer models. They used ad-hoc estimates for the turbidity parameters (aerosol optical depth and precipitable water) which tended more toward the climatic norm. Turbidity data have not been widely available nor existed for long periods of record. Hence, they could not utilize the annual trends developed from 30 year climatic lows in broadband aerosol optical depth and precipitable water. Resulting clear-day models tend to under-estimate the climatic boundary for maximum data selected from a 30 year span. A better model would eliminate or reduce this problem. In addition, it could also serve as a dynamic upper boundary for screening global solar radiation data.

The recently developed Solar and Meteorological Observation Network database (SAMSON) provides, as recommended by the WMO (1967) standards, a 30 year serially-complete and quality-controlled record of hourly and daily global solar radiation data and many other meteorological variables, including the previously stated turbidity parameters, for hundreds of locations throughout the United States. Hence for locations within the United States it is possible to develop and test a radiative-transfer model using appropriate climatic extreme data, maxima for global solar radiation and minima for turbidity variables, provided the turbidity parameters are accurate in time and space and the local surface albedo is known. Long term studies of variability in turbidity are rare, especially for locations that correspond to a site in the SAMSON database; Zymber and Sellers (1985) may be the only such report and it tables 27 years of Linke turbidity factor for Tucson, AZ but only on a monthly average basis.

4. Areal scales of fluctuation for determining averaging areas

Point-source surface meteorological data from individual stations and from a network of stations, like those within any one of the many recent field experiments, are the basis of development, comparison, and scaling up for areal and regional estimates of evapotranspiration (e.g., see Kustas
and Goodrich 1994). In many of the field scale experiments and in all the
data sets included in the recently released ISLSCP (the International
Satellite Land Surface Climatology Project) Global Data sets very simple
spatial aggregation techniques were used (Sellers et al. 1996). It is
well-known that land surfaces can be heterogeneous in space and time (see
e.g., Chen and Coughenour 1994). This heterogeneity may be the reason, at
least in part, why remote-sensing data for such hydrological studies have
not been helpful (see e.g., Kirby 1996; Moran et al. 1994; Bolle 1993b; or
Brutsaert and Sugita 1992). Recall, as previously stated, that many
researchers believe the need to reformulate processes at appropriate scales
is primary to regional hydrological modeling. To do such a job it is
necessary to understand the variability of the basic parameters and
processes (e.g., see Chapter 6 in Lin and Segel 1974). Some related
theoretical work on evapotranspiration and energy balance has been done.
Gash (1987) presented an error analysis for energy balance on real-time
spatial data. Raupach (1993) discussed aggregation techniques based on
convective boundary-layer theory assuming the underlying spatial
variability of the surface is already characterized. Theoretical
approaches for temporally heterogeneous systems have been given by Rose
(1984) and yet static vegetation land-surface models are used in all
general circulation climate models (Chen and Coughenour 1994). Although
there have been geostatistical studies on long-term reference
evapotranspiration estimated with monthly or longer-based statistical
evapotranspiration models (Martínez-Cob 1996; Martínez-Cob and Cuenca 1992;
Cuenca and Amegee 1987), no-one has directly addressed the spatial and
temporal scales of variability in the energy-balance data or related
surface parameters with spatial and time-series statistics. Some of the
remote-sensing data associated with field experiments, the ISLSCP data
sets, and EDC (the USGS [United States Geological Survey] ERDOS Data
Center) data sets offer an opportunity to do formal spatial and temporal
analyses. Relevantly, a formal methodology for this job exists, the scale
of fluctuation, and has been argued for by Rodriguez-Iturbe (1986). It is based on Vanmarcke's random-field theory (1983). More recently Raupach (1993) stressed the need for defining an aggregation area but did not offer a method. In fact, although Rodriguez-Iturbe's (1986) paper was based on a simulation study, he called for case studies based on real data. Spatial albedo data, as previously argued, is an important surface property for net radiation and the energy balance. Moreover, although there are problems, such data can be constructed from AVHRR data (e.g., see Bolle 1993b or Ranson et al. 1991) to conduct a real-data based case study.
DATA QUALITY CHECKING FOR SINGLE STATION METEOROLOGICAL DATABASES\textsuperscript{1}

A paper published in Agricultural and Forest Meteorology\textsuperscript{2}

D.W. Meek\textsuperscript{3} and J.L. Hatfield\textsuperscript{4}

Abstract

In the past decade individual and networks of automated meteorological stations have been installed throughout the United States and many other countries. For a variety of reasons, the data collected are being archived in databases; however, quality control/quality assurance procedures, when employed, vary greatly. As a start to possible standardization, screening rules for hourly and daily data values are proposed for quality checking micrometeorological data from individual base stations that record solar irradiance (SI), precipitation (P), barometric pressure (P\textsubscript{b}), vapor pressure (e), wind speed (u\textsubscript{10}), wind direction (θ\textsubscript{10}), air temperature (T\textsubscript{a}), and three soil temperatures (T\textsubscript{s1}, T\textsubscript{s2}, and T\textsubscript{s3}). Three types of screening rules are considered: (1) high/low range limits (LIM), (2) rate-of-change limits (ROC), (3) continuous no-observed-change with time limits (NOC). Daily data from historical meteorological records for Ames, IA (30 y) and Treynor, IA (26 y) were available for developing climatic based dynamic data screening rules. Otherwise, instrument

\textsuperscript{1}Contribution of the Agricultural Research Service, U.S. Department of Agriculture. (Received 1 March 1993; revision accepted 29 October 1993)


\textsuperscript{3}Mathematician, United States Department of Agriculture, Agricultural Research Service-Midwest Area, National Soil Tilth Laboratory, 2150 Pammel Drive, Ames, IA 50011-4420, USA. Also corresponding author.

\textsuperscript{4}Laboratory Director, United States Department of Agriculture, Agricultural Research Service-Midwest Area, National Soil Tilth Laboratory, 2150 Pammel Drive, Ames, IA 50011-4420, USA. Also Professor (USDA Collaborator), Department of Agronomy, Iowa State University.
specifications and theoretical models were used to develop screening rules for the remaining measurements. Hourly and daily data from well maintained, automated weather stations at Walnut Creek, IA (9 mon) and Treynor, IA (1 y) were used to evaluate and refine the screening rules. Daily data are not flagged often. The most common flag, on either time scale, was on vapor pressure when its value exceeded the 95% relative humidity calibration limit of the sensor. Hourly SI often exceeded a computed extraterrestrial radiation value, particularly at sunset. Rule 1 (LIM) is mainly invoked via observations outside the sensor ranges; rule 2 (ROC) flags abrupt changes; rule 3 (NOC) flags unusually steady periods in the data stream. When used as part of a total field operation and data processing system, these rules improve the data quality and may help with data exploration.

1. Introduction

For more than a decade many of the agricultural producer states, particularly California and Nebraska (Meyer and Hubbard, 1992; Snyder and Pruitt, 1992), have set up networks of individual meteorological base stations similar to those described by Howell et al. (1984). These stations include solar radiation, precipitation, vapor pressure, barometric pressure, wind speed, wind direction, air temperature, and soil profile temperature sensors (Hubbard et al., 1983). For purposes of scientific research and resource management, particularly in agriculture, hourly and daily data from these stations are being put into databases. All such efforts are enhanced by quality checking the data; however, no consistent methodology is employed. For example, Ashcroft et al. (1990a, b), employ graphical display and human judgment, while the High Plains Climate Center
employs a fixed acceptable range for variable and regional maps (Kenneth Hubbard, personal communication, 1991). Perhaps, as a consequence of non-existing or mixed quality control methods, in some agronomic research involving models that use meteorological data, questionable results have been attributed to poor data quality (Hatfield and Fuchs, 1990). Hence, there is a need to set and standardize data quality assurance/quality control (QA/QC) methods.

Quality assurance/quality control for meteorological data involves several steps. These are divided into field operations and data processing operations. Field operations start with setting instrument specifications and calibration procedures. Field installation, routine and special maintenance, as well as periodic field instrument checks, follow as the next step. The latter utilizes QA/QC procedures that can include statistical methods and software. Howell et al. (1984) presents practical operational guidelines in greater detail. Data processing procedures are an additional layer after field procedures. If a station is part of a network, data processing procedures can be either stand-alone or network based. If soundly performed with given specific objectives and methods, network-based analyses are extremely useful. In most cases, such procedures are empirically derived for specific locations and purposes. Moreover, they require a reasonable length of record with good quality data as well as development time and expertise. Single stations, new stations, and new networks will not be able to be checked with such procedures. Hence, we developed stand-alone data processing procedures only as a distinct part of the total QA/QC process. Network analyses can be added on to a software system to provide additional checking of multiple stations.
For the data processing of stand-alone measurements, a methodological precedent is available. Using real-time hydrological data, O'Brien and Keefer (1985) proposed a set of three computer based rules that eliminate human intervention, but do not require a more thorough set of analyses; they are: (1) fixed or dynamic high/low bounds for each variable (LIM); (2) fixed or dynamic rate-of-change limits for each variable (ROC); (3) a continual no-observed-change in time limit (NOC).

O'Brien and Keefer (1985) developed and validated these rules on stand-alone stream gaging data but not on meteorological data. Such rules also seem highly appropriate for most stand-alone meteorological station data. The ROC and NOC rules can be thought of as an upper and lower bound on the rate of change for a given variable. This paper develops appropriate data screening rule versions from daily historical meteorological data, instrument specifications, and other physical considerations.

2. Materials and methods

Table 1 lists geographic and length of record information about the sites and data. All sites are located within fenced areas at least 25 m away from any building or large natural body. The development sites are on short clipped grass except for the soil temperature thermometers which are in bare soil. The rule test sites are on native vegetation. The development sites have the air temperature thermometers in white-painted wooden "Cotton Belt" shelters (United States Department of Commerce, National Weather Service [USDC-NWS], 1972). The rule test sites have the automated instruments configured on or around a tripod mast (see e.g., Howell et al., 1984). All sites are within Iowa, which has a continental
temperate climate. The historical data from Ames and Treynor were used to develop the screening rules. The Ames data are from the USDC-NWS records (1960-1990). The Treynor data are from the Deep Loess Research Station daily records (United States Department of Agriculture, Agricultural Research Service [USDA-ARS]).

Installation and operation of the automated weather stations was similar to that outlined in Howell et al. (1984). Factory calibrations for each instrument were used. Installation and set-up followed the instrument and micrologger manuals. Suggested daily, weekly, and longer term checks, maintenance, and set instrument replacements were scheduled. Maintenance logs for each station were kept. Table 2 details the instrumentation used and measurements available at each site.

The screening rules were developed and subsequently tested on the data sets from the two automated stations. The bounds for LIM rules are based either on static or dynamic climatic extremes or on the response ranges of the given sensors. The upper bounds for ROC rules are based on either climatic extremes when possible or diminishing fractions of the response range generally starting with one half of the response range. The NOC rules always use a minimum of two lag values. Unavoidably, the task of setting all such bounds is somewhat arbitrary. We emphasize the approach rather than the absolute limits which can be adjusted according to user needs. Hourly and daily results are presented separately because many users are only interested in daily data. The complete set of the screening rules is listed in Tables 3-6.
3. Rule development for daily data records

3.1 Solar irradiance rules

Because solar irradiance (SI) data were available for the Ames site, only one set of daily SI rules was developed, although we tested them at both locations. The LIM rule for daily data bounds the data from above with a sinusoidal clear day curve, \( SI_c(d) \) (Howell et al., 1984), and below with zero. The clear day curve was developed using the maximum daily value taken for each day in the year over the 31 yr period of data collection and regressing eqn. (1), the fundamental term in a trigonometric Fourier series, on the resulting set. The parameters \( a_0 \) to \( a_2 \) are the regression estimates; \( d \) represents day of the year. Formally, the daily SI LIM rule is given by the following inequality (in MJ m\(^{-2}\))

\[
0 \leq SI(d) \leq SI_c(d)
\]

The proposed ROC rule for daily SI data is

\[
0 \leq |SI(d) - SI(d-1)| \leq \Delta SI(d)
\]

The upper bound ROC function, \( \Delta SI(d) \), was developed from regressing a cosine curve with the same parameter form as eqn. (1), to a data set generated by taking the maximum value of

\[|(SI(d) - SI(d-1))|\]

from 1 day for each day in the year over the 31 yr period of data collection.

The NOC rule for daily SI is:

\[\neg SI(d) = SI(d-1) = SI(d-2)\]

The symbol "\( \neg \)" is a logical negation sign.
3.2 Air temperature rules

Daily $T_a$ extremes were available for both locations. The LIM rule for daily data bounds the data from above and below with sinusoidal maximum and minimum curves denoted $T_{a\, \text{max}}(d)$ and $T_{a\, \text{min}}(d)$, respectively. Formally, the daily $T_a$ LIM rule is given by the following inequality (in °C)

$$T_{a\, \text{min}}(d) \leq T_a(d) \leq T_{a\, \text{max}}(d)$$

These curves were developed using the maximum or minimum daily value taken for each day in the year over the period of data collection at each location and regressing an equation of a form similar to eqn. (1) on each individual site-extreme data set.

The proposed ROC rule for daily $T_a$ data is (in °C)

$$0 \leq |T_a(d)-T_a(d-1)| \leq \Delta T_a(d)$$

Upper bounds for the ROC screening function for both locations, $\Delta T_a(d)$, were developed from regressing a sine curve, similar in form to eqn. (1), to each set generated by taking the maximum value of $|T_a(d)-T_a(d-1)|$ for each day in the year over the period of data collection at each location.

The NOC rule for daily $T_a$ is

$$-T_a(d)=T_a(d-1)=T_a(d-2)$$

3.3 Soil temperature rules

Monthly $T_s$ extremes at approximately 0.010 and 0.020 m were available only for the Ames location. Similar to $T_a$, the LIM rule for soil data bounds the data from above and below with sinusoidal maximum and minimum curves. Formally, the daily $T_s$ LIM rules are given by the following inequalities (in °C)

$$T_{s10\, \text{min}}(d) \leq T_{s10}(d) \leq T_{s10\, \text{max}}(d)$$

and
These curves were developed using the maximum and minimum monthly value taken from one day for each mid-month date in the year over the 31 yr period of data collection at the Ames location and regressing an equation of a form similar to eqn. (1) on each individual depth-extreme data set.

Because the soil temperature data used were not on a daily basis, the upper bounds on the respective ROC rules are constants; formally for each depth, they are

\[ 0 \leq |T_{s10}(d) - T_{s10}(d-1)| \leq 2.5^\circ C \]

and

\[ 0 \leq |T_{s20}(d) - T_{s20}(d-1)| \leq 2.0^\circ C. \]

The NOC rules for daily \( T_s \) at each depth are as follows

\[ T_{s10}(d) = T_{s10}(d-1) = T_{s10}(d-2) \]

and

\[ T_{s20}(d) = T_{s20}(d-1) = T_{s20}(d-2) \]

At both sites there is a soil temperature probe near the soil surface, \( T_s \).

The air temperature rules are used to screen these data because there are limited historical data available for this measurement.

3.4 Precipitation rule

Precipitation events exhibit a great deal of randomness, so simple static rules are employed. The LIM rule for daily data, \( P(d) \), bounds the data from above with a constant value and below with zero. Formally, the daily \( P \) LIM rule is given by the following inequality (in mm):

\[ 0 \leq P(d) \leq P_{\text{max}}. \]

The \( P_{\text{max}} \) for each site was arbitrarily selected by rounding up to the next 10 mm level the maximum daily value taken from the entire period of data.
collection for each corresponding location. There are no proposed ROC and NOC rules because in our region single storm events lasting more than a day are very rare.

3.5 Vapor pressure rules

For vapor pressure and for the remaining types of daily data, simple static rules are employed. The LIM rule for daily data, \( e(d) \), transforms the datum to the corresponding relative humidity then bounds the data from both above and below with constant values. The saturation vapor pressure \( (e_s(d)) \), used to calculate relative humidity, is estimated with the Goff-Gratch Equation (Buck, 1981) using average daily air temperature, \( T_a(d) \).

Formally, the data \( e \) LIM rule is given by the following inequality:

\[ 0.15 \leq \frac{e(d)}{e_s(d)} < 0.96 \]

The limiting values were selected based on the listed calibration of the sensors employed at the test sites (Campbell Scientific, 1990c).

The proposed ROC rule for daily \( e \) data is: \( 0 \leq |(e(d)-e(d-1)| \leq 2 \) kPa. The upper bound for the ROC screening is the same value for both test locations. The NOC rule for daily \( e \) is once more a two lag rule

\[-e(d)=e(d-1)=e(d-2)\]

3.6 Barometric pressure rules

The LIM rule for daily \( P_b(d) \) bounds the data from both above and below with constant values. Formally, the LIM rule is:

\[ 88.0 \leq P_b(d) \leq 106.0 \text{ kPa} \]

The upper limiting value was selected based on the listed calibration of the sensor employed at both test sites (Campbell Scientific, 1992). Because the lower limit of the sensor, 80 kPa, is unusually low for
realistic readings in lower elevations, a higher value was chosen based on record lows at sea level in the eye of hurricanes (Wallace and Hobbs, 1977). While there are no normal pressure data available for either site, theoretical estimates are possible; the expectation values for the Walnut Creek and Treynor test sites are 97.5 kPa and 96.6 kPa, respectively. The estimates are based on an exponential correction to standard sea level pressure, 101.3 kPa; the model uses each site elevation and assumes an 8.5 km scale height for the atmosphere (Wallace and Hobbs, 1977).

The daily based ROC rule for both sites is
\[ 0 \leq |(P_b(d) - P_b(d-1))| \leq 4 \text{ kPa} \]
The daily based NOC rule for both sites is
\[ -P_b(d) = P_b(d-1) = P_b(d-2) \]

3.7 Wind speed rules

The LIM rule for daily wind speed, \( u_2(d) \), bounds the data from both above and below with constant values taken from the instrument specifications (Campbell Scientific, 1990a). Formally, the LIM rule is
\[ 0.45 < u_2(d) < 45.00 \text{ ms}^{-1} \]
Notice that the actual bounds are excluded.

The daily based ROC rule for both sites is
\[ 0 \leq |(u_2(d) - u_2(d-1))| < 10 \text{ ms}^{-1} \]
The daily based NOC rule for both sites is
\[ -u_2(d) = u_2(d-1) = u_2(d-2) \]

3.8 Wind direction rules

The LIM rule for daily wind direction, \( \theta_2(d) \), bounds the data from both above and below with constant values taken from the instrument manual
Formally, the LIM rule is

\[ 0 \leq \theta_1(d) < 360^\circ \text{ compass} \]

The daily based ROC rule for both sites is

\[ 0 \leq \min(|\theta_1(d) - \theta_1(d-1)|, 360 - |\theta_1(d) - \theta_1(d-1)|) < 150^\circ \text{ compass} \]

The ROC rule excludes a reversal in wind direction. The daily based NOC rule for both sites is

\[ -\theta_1(d) = \theta_2(d-1) = \theta_2(d-2) \]

4. Rule development for hourly data records

Long-term records on hourly data for all parameters were not available, so proposed bounds were estimated from theoretical considerations, instrument specifications, or modifications of daily rules. In general, the hourly rules are similar to the daily ones with the ROC rule bounded with a smaller constant value and the number of lags in the NOC rule increased to a minimum of three. Except as noted, the rules apply to both test sites.

4.1 Solar radiation rules

The LIM rule for hourly SI data is [in MJ m\(^{-2}\)]

\[ 0 \leq SI(d,h) \leq SI_{\text{ext}}(d,h) \]

with the upper bound for the data being the computed extraterrestrial radiation value; here Iqbal's model (Iqbal, 1983) is employed (the latitude and longitude of each site are required input). Because the SI instrument used at the test sites is not a high precision sensor (LI-COR, 1986), modeling the atmospheric transmission of shortwave irradiance was deemed
unnecessary for our purposes. The ROC rule for hourly SI data is [in MJ m\(^{-2}\)]

\[ 0 \leq |\text{SI}(h) - \text{SI}(h-1)| \leq 2 \]

The NOC rule for hourly SI is

\[ \sim \text{SI}(h) = \text{SI}(h-1) = \text{SI}(h-2) = \text{SI}(h-3) \]

only when SI(h) is greater than 0.

4.2 Air temperature rules

The hourly \( T_a \) \(_{\text{LIM}} \) rule is a modification of the daily rule. Formally, it is

\[ T_a^{\text{max}}(d) - 2.5 \leq T_a(h) \leq T_a^{\text{max}}(d) + 2.5°C \]

Notice that the hourly temperature range is allowed to exceed the daily. While these limits are arbitrary and less conservative, there are two arguments for the practice. Firstly, the daily extreme limits were developed from data acquired from liquid in glass recording thermometers while the automated data were acquired from thermistor measurements recorded at one-minute intervals. In principle, the latter respond more rapidly and so can capture shorter duration extreme observations. Secondly, the regression curve for each extreme averages the seasonal trend in the respective daily extreme. There are lack-of-fit, standard error of the estimate, and other measures of imperfection associated with each curve. The expanded range helps mask out small scale variation near the extremes. The hourly ROC rule is

\[ 0 \leq |T_a(h) - T_a(h-1)| \leq 6°C \]

The hourly NOC rule for \( T_a \) is

\[ \sim T_a(h) = T_a(h-1) = T_a(h-2) = T_a(h-3) \]
4.3 Soil temperature rules

The hourly $T_s$ LIM rules for the three depths are given by the following inequalities:

- $T_{s\min}(h) \leq T_{s0}(h) \leq T_{s\max}(h)$
- $T_{s10\min}(h) \leq T_{s10}(h) \leq T_{s10\max}(h)$
- $T_{s20\min}(h) \leq T_{s20}(h) \leq T_{s20\max}(h)$

The upper bounds on the respective hourly ROC rules are lower than the daily values. The ROC rules are as follows.

- $0 \leq |T_{s0}(h) - T_{s0}(h-1)| < 2.5^\circ C$
- $0 \leq |T_{s10}(h) - T_{s10}(h-1)| < 1.0^\circ C$
- $0 \leq |T_{s20}(h) - T_{s20}(h-1)| < 1.0^\circ C$

At below surface depths, $T_s$ can be steady, so for the hourly NOC rules four lags are proposed. The NOC rules are as follows:

- $T_{s0}(h) = T_{s0}(h-1) = T_{s0}(h-2) = T_{s0}(h-3)$
- $T_{s10}(h) = T_{s10}(h-1) = T_{s10}(h-2) = T_{s10}(h-3) = T_{s10}(h-4)$
- $T_{s20}(h) = T_{s20}(h-1) = T_{s20}(h-2) = T_{s20}(h-3) = T_{s20}(h-4)$

4.4 Precipitation rules

The LIM rule is (in mm) $0 \leq P(h) \leq P_{\max}$. Again, the $P_{\max}$ value for both sites was set by estimating the volume corresponding to the maximum number of tips the instrument can sustain without resulting in an error. The ROC rule for hourly $P$ is

- $0 \leq |P(h) - P(h-1)| \leq \frac{1}{2}P_{\max}$

It is only used when $P(h)$ is greater than 0. The NOC rule for hourly $P$ is

- $P(h) = P(h-1) = P(h-2) = P(h-3)$

It is also only used when $P(h)$ is greater than 0.
4.5 Vapor pressure rules

The hourly LIM rule is the same as the daily rule

\[ 0.15 \leq \frac{\text{e}(h)}{\text{e}(h)} < 0.96 \]

The upper limit for the ROC rule for e data is lower

\[ 0 \leq |\text{e}(h) - \text{e}(h-1)| \leq 1 \text{ kPa} \]

The NOC rule for hourly e uses a three lag rule

\[ \text{e}(h) = \text{e}(h-1) = \text{e}(h-2) = \text{e}(h-3) \]

4.6 Barometric pressure rules

The LIM rule is the same as the daily one

\[ 88.0 \leq \text{P}_b(h) \leq 106.0 \text{ kPa} \]

The ROC rule is

\[ 0 \leq |\text{P}_b(h) - \text{P}_b(h-1)| \leq 1 \text{ kPa} \]

The NOC rule changes. It uses 11 lag values because the pressure distribution can on occasion remain stationary for long periods; hence, the NOC rule is

\[ \text{P}_b(h) = \text{P}_b(h-1) = \text{P}_b(h-2) = \ldots = \text{P}_b(h-11) \]

4.7 Wind speed rules

The LIM rule for hourly wind speed is the same as the daily

\[ 0.45 < u_2(h) < 45.00 \text{ ms}^{-1} \]

The hourly ROC rule has a smaller upper bound

\[ 0 \leq |(u_2(h) - u_2(h-1))| < 10 \text{ ms}^{-1} \]

The NOC rule is

\[ u_2(h) = u_2(h-1) = u_2(h-2) = u_2(h-3) \]
4.8 Wind direction rules

The LIM rule for hourly wind direction is the same as the daily one

\[ 0 \leq \theta_2(h) \leq 360^\circ \text{ compass} \]

The hourly ROC rule is also the same as the daily one

\[ 0 \leq \min (|\theta_2(h) - \theta_2(h-1)|, 360 - |\theta_2(h) - \theta_2(h-1)|) < 150^\circ \text{ compass} \]

The NOC rule uses three lag values

\[-\theta_2(h) = \theta_2(h-1) = \theta_2(h-2) = \theta_2(h-3)\]

5. Data processing

In our database, data flags are stored as character strings while data are stored as numbers. While this is not the most compact storage method, it makes access easy. If a datum is flagged by one of the screening rules, a single flag character is stored instead of a blank. We use the following character scheme for screening rule violations: (1) For the LIM rule, H or L for high and low flag; (2) For the ROC rule, D for difference flag; (3) For the NOC rule, C for constant flag.

6. Results

6.1 Daily data records

A summary of the daily rules along with results for each rule at each test site is shown in Tables 3 and 4 for Walnut Creek and Treynor, respectively. Fig. 1 shows the LIM and ROC rules for \( T_s(d) \) at Walnut Creek. Notice the only outlier there, Day 311, at the end of a sequence of dropping values; it is a valid observation! Starting on Day 306, a rain storm became more severe, most of the state area experienced a heavy snow storm that brought normal daily activities to a halt. The unusual cold spell lasted over a week. Many of the flagged data are associated with
this event. The sole flagged P_b value was on that day with ROC=6 kPa. The
P_b pattern at Treynor was similar, but the drop was not as abrupt. Based
on the hourly u_t values, the daily average value was set to missing for 5
days in a row because the sensor appears to have frozen during several
subintervals within this period. We base this assessment on several
related facts: (1) the air temperature was below 0°C; (2) the vapor
pressure was at or near saturation; (3) the wind direction data showed
changes; and (4) ice was observed on the tripod and instruments. It should
be noted, however, the deletion of observations here was only for the
assessment of the screening rules. In practice, no data should be deleted
but just flagged.

Other than abrupt or severe events, flagged observations were rare on
a daily basis. The one possible exception, which is also the most
frequently occurring flag in our climate, (= 7% of time) is an upper LIM
rule violation on e(\delta). For practical purposes, however, the air could be
considered saturated at these times.

6.2 Hourly data records

A summary of the hourly rates along with results for each rule at each
test site is shown in Tables 5 and 6 for Walnut Creek and Treynor,
respectively. Fig. 2 shows the LIM rule residuals vs. h for SI(h) at
Walnut Creek. Notice that at both sites about 7% of the SI(h) values are
flagged by the LIM rule. These are mostly occurring at sunset hours.
There are also a few questionable sunrise values but they are not as
pronounced. Some possible reasons are as follows: (1) the radiometer is
not a high precision instrument (LI-COR, 1986); (2) it does not sense the
full shortwave band (ibid); (3) it has poor cosine angle response at low
angles (ibid); (4) it has no temperature correction (ibid); (5) there may be reflections from clouds; (6) the datalogger time may be incorrect. For practical purposes, however, these observations could be considered zero because the observations are generally small. Given these conditions, we think that using an atmospheric transmission model for the upper LIM is unwarranted.

As with the daily data, high $e(h)$ values also occur often, in fact, more frequently. Again, via the ROC test, abrupt or severe events were readily discernible in most parameters. In contrast, NOC flagged values revealed several periods of steady conditions in $e(h)$, $T_{h}(h)$, $T_{v}(h)$, and $P_{a}(h)$; the latter two often had sequences of flagged values. Otherwise, except as noted, flagged observations were generally unusual.

7. Discussion

Data quality assurance is only partially achieved with stand-alone screening rules. As previously stated, there must be an effective field operations program as well as a software system. Instrument calibration drifts, slow instrument failures, or incidences of temporary monitoring interference are difficult to detect with data analyses alone. Moreover, more thorough data analyses for all of the data would consume a lot of human and computer resources, yet the results would still be speculative. Users who require a more thorough screening of the data are probably few and far between. Also, such users would probably choose to implement their own screening methods anyway. Such an effort would be on the level of an expert system.

When implementing the screening rules, we recommend using them hierarchically. If a datum is flagged by the LIM rule, other possible
problems are probably obvious and not as interesting. Consider the frozen wind speed sensor on hourly data. With the proposed rules, the first three observations would be captured by the LIM while the fourth would be captured by both the LIM and NOC rules. Because the LIM rule is violated, the datum is already suspect and the reason for the NOC violation is clear. Also, this system allows us to keep the single character flag scheme. In contrast, the reporting of a multiple violation in the data storage would involve a more complicated flagging system. At the first observation or at datum following a missing value, the datum is flagged with an 'S'. The reason for noting such observations is that the ROC and NOC rules cannot be implemented. Finally, daily values derived from flagged hourly values should also be flagged; we denote this with an 'F'.

Another important issue in implementing a database is that of handling missing and flagged values. Some of the developing state micrometeorological databases (e.g., Hubbard et al., 1983; Snyder and Pruitt, 1992) provide missing data values based on empirical relations derived with data from other nearby stations. Some modelers prefer to use their own methods, appropriate for their goals, to estimate missing values. When observations are missing in our data, we will leave them as missing. We believe the estimation of missing data is a subject unto itself which needs to be addressed. Similarly, we believe that the use or rejection of a flagged value, especially LIM violations, is the responsibility of the user. For example, for practical purposes, an e(h) or e(d) value flagged with H could be replaced with a modeled e, value.

Finally, there are some operational and long-term considerations to ponder. If the database is to be collected and maintained for a long
period, say over 10 years, then the screening rules based on the climatic
data should probably be recalibrated periodically. Some researchers or
users may prefer to transform or alter the measured variables and so
appropriately change screening rules; for example, using dew point instead
of vapor pressure for the moisture measurement. Another is extending the
wind direction scale from 0 to 540° or using wind vectors because period
averages during which the direction crossed 0° (i.e. from northerly
directions) are incorrect. Lastly, as the microloggers on the automated
weather stations become more sophisticated, the screening procedures and
flagging system could be implemented within the micrologger software.

8 Conclusions

The LIM, ROC, and NOC rules should be used hierarchically. The LIM
rule is invoked when observations exceed either sensor limits or
climatically determined local extremes. The ROC rule detects abrupt
changes in the data stream. The NOC rule shows unusually steady periods in
the data stream. Used along with regular field maintenance and quality
control procedures, these rules can help ensure data quality. Although
there are many other ways to explore and analyze data, the flagging system
can also serve as a selection tool for this purpose, particularly in
finding real but unusual events.

Acknowledgments

The authors would like to thank all the following: Professor R.E.
Carlson, ISU, Ames, IA, for comments on the manuscript and for the Ames
data file; the staff of the ISU Agronomy Farm, Boone, IA, for collection
and recording of the Ames historical data; L.A. Kramer and the staff of the
USDA-ARS-MWA-DLRS, Treynor, IA, for collecting and recording the Treynor data; Dr. K.G. Hubbard, UNL, Lincoln, NE, for comments on the manuscript and for providing information on Nebraska's weather data system; G. McCurdy, State Climatology Office, USU, Logan, UT, for providing information on Utah's weather data system; Dr. T.A. Howell, USDA-ARS-SPA-WMR, Bushland, TX, and Dr. J. Prueger, USDA-ARS-MWA-NSTL, Ames, IA, and the reviewers for their comments on the manuscript.

References


Table 1

Geographic and historic information for the meteorological stations

<table>
<thead>
<tr>
<th>Location name</th>
<th>Northern latitude</th>
<th>Western longitude</th>
<th>Elevation</th>
<th>Period of record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule development sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ames, IA</td>
<td>42° 2.0'</td>
<td>93° 48.0'</td>
<td>335.0 m</td>
<td>1960 - 1990</td>
</tr>
<tr>
<td>Treynor, IA</td>
<td>41° 12.0'</td>
<td>95° 38.5'</td>
<td>378.0 m</td>
<td>1964 - 1990</td>
</tr>
<tr>
<td>Rule test sites</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walnut Creek, IA</td>
<td>41° 57.5'</td>
<td>93° 38.3'</td>
<td>304.8 m</td>
<td>278 days in 1991</td>
</tr>
<tr>
<td>Treynor, IA</td>
<td>41° 12.0'</td>
<td>95° 38.5'</td>
<td>378.0 m</td>
<td>365 days in 1991</td>
</tr>
</tbody>
</table>
## Table 2
Measurements recorded at each location

<table>
<thead>
<tr>
<th>Location</th>
<th>Parameter</th>
<th>Symbol</th>
<th>Instrument*</th>
<th>Storage frequency</th>
<th>Observational frequency</th>
<th>Units used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rule development sets</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ames, IA</td>
<td>Total solar radiation</td>
<td>SI(d)</td>
<td>Eppley PSP</td>
<td>d^4</td>
<td>Continuous^</td>
<td>M J m^2</td>
</tr>
<tr>
<td></td>
<td>Maximum air temperature</td>
<td>T_a in (d)</td>
<td>Liquid &amp; Glass Max/Min</td>
<td>d^4</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Minimum air temperature</td>
<td>T_n in (d)</td>
<td>Liquid &amp; Glass Max/Min</td>
<td>d^4</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Total precipitation</td>
<td>P(d)</td>
<td>8 inch gage</td>
<td>d^4</td>
<td>d^4</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Maximum soil temperature, 4&quot;</td>
<td>T_11 in (d)</td>
<td>Palmer Soil</td>
<td>mon^1</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Minimum soil temperature, 4&quot;</td>
<td>T_21 in (d)</td>
<td>Palmer Soil</td>
<td>mon^1</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Maximum soil temperature, 8&quot;</td>
<td>T_12 in (d)</td>
<td>Palmer Soil</td>
<td>mon^1</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Minimum soil temperature, 8&quot;</td>
<td>T_22 in (d)</td>
<td>Palmer Soil</td>
<td>mon^1</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td>Treynor, IA</td>
<td>Maximum air temperature</td>
<td>T_a in (d)</td>
<td>Liquid &amp; Glass Max/Min</td>
<td>d^4</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Minimum air temperature</td>
<td>T_n in (d)</td>
<td>Liquid &amp; Glass Max/Min</td>
<td>d^4</td>
<td>d^4</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Total precipitation</td>
<td>P(d)</td>
<td>8 inch gage</td>
<td>d^4</td>
<td>d^4</td>
<td>mm</td>
</tr>
<tr>
<td><strong>Rule test sets</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>Total solar radiation</td>
<td>SI(h)</td>
<td>Licor 200S</td>
<td>h^4</td>
<td>min^3</td>
<td>M J m^2</td>
</tr>
<tr>
<td></td>
<td>Total solar radiation</td>
<td>SI(d)</td>
<td>Licor 200S</td>
<td>d^4</td>
<td>min^3</td>
<td>M J m^2</td>
</tr>
<tr>
<td></td>
<td>Average air temperature</td>
<td>T_a(h)</td>
<td>Campbell HMP35C</td>
<td>h^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average air temperature</td>
<td>T_n(h)</td>
<td>Campbell HMP35C</td>
<td>d^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 40 mm</td>
<td>T_p(h)</td>
<td>Campbell 107</td>
<td>h^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 40 mm</td>
<td>T_p(d)</td>
<td>Campbell 107</td>
<td>d^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 100 mm</td>
<td>T_1(h)</td>
<td>Campbell 107</td>
<td>h^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 100 mm</td>
<td>T_1(d)</td>
<td>Campbell 107</td>
<td>d^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 200 mm</td>
<td>T_2(h)</td>
<td>Campbell 107</td>
<td>h^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Average soil temp., 200 mm</td>
<td>T_2(d)</td>
<td>Campbell 107</td>
<td>d^4</td>
<td>min^3</td>
<td>°C</td>
</tr>
<tr>
<td></td>
<td>Total precipitation</td>
<td>P(h)</td>
<td>Campbell Te525</td>
<td>h^4</td>
<td>min^3</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Total precipitation</td>
<td>P(d)</td>
<td>Campbell Te525</td>
<td>d^4</td>
<td>min^3</td>
<td>mm</td>
</tr>
<tr>
<td></td>
<td>Average vapor pressure</td>
<td>e(h)</td>
<td>Campbell HMP35C</td>
<td>h^4</td>
<td>min^3</td>
<td>kPa</td>
</tr>
<tr>
<td></td>
<td>Average vapor pressure</td>
<td>e(d)</td>
<td>Campbell HMP35C</td>
<td>d^4</td>
<td>min^3</td>
<td>kPa</td>
</tr>
<tr>
<td></td>
<td>Average barometric pressure</td>
<td>P_h(h)</td>
<td>Vaisala PTA427</td>
<td>h^4</td>
<td>min^3</td>
<td>kPa</td>
</tr>
<tr>
<td></td>
<td>Average barometric pressure</td>
<td>P_h(d)</td>
<td>Vaisala PTA427</td>
<td>d^4</td>
<td>min^3</td>
<td>kPa</td>
</tr>
<tr>
<td></td>
<td>Average windspeed</td>
<td>u_h(h)</td>
<td>Met-One 014A</td>
<td>h^4</td>
<td>min^3</td>
<td>m s^-1</td>
</tr>
<tr>
<td></td>
<td>Average windspeed</td>
<td>u_h(d)</td>
<td>Met-One 014A</td>
<td>d^4</td>
<td>min^3</td>
<td>m s^-1</td>
</tr>
<tr>
<td></td>
<td>Average wind direction</td>
<td>θ(h)</td>
<td>Met-One 024A</td>
<td>h^4</td>
<td>min^3</td>
<td>°Compass</td>
</tr>
<tr>
<td></td>
<td>Average wind direction</td>
<td>θ(d)</td>
<td>Met-One 024A</td>
<td>d^4</td>
<td>min^3</td>
<td>°Compass</td>
</tr>
</tbody>
</table>

* The mention of a trade name does not imply endorsement by the USDA-Agricultural Research Service.

^ The letters SI stand for shortwave irradiance, i.e. the bandwidth the instrument is measuring.

° SI, u, and θ are measured at 2 m elevation above the ground. T_p, P, and e are measured at 1.5 m above the ground. The soil temperatures T_1, T_2, and T_3 are at the depths indicated.

^ d, day; mon, month; h, hour.

° Strip chart recording of PSP voltage output, integrated into daily irradiances.
### Table 3
Tests on screening rules for Walnut Creek daily data

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Symbol</th>
<th>Units</th>
<th>Rule name</th>
<th>Condition for acceptable data</th>
<th>Test set</th>
<th>Total no. obs.</th>
<th>No. excluded (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar irradiance</td>
<td>SI(d)</td>
<td>MJ m⁻²</td>
<td>LIM</td>
<td>0 ≤ SI(d) &lt; SI(d)⁺</td>
<td>285</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air temperature</td>
<td>T_d</td>
<td>°C</td>
<td>LIM</td>
<td>Tₚₙₙ(d) ≤ Tₖ(d) ≤ Tₚₙₙ(d)</td>
<td>283</td>
<td>1</td>
<td>(0.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 4 cm</td>
<td>Tₚₚₙₙ(d)</td>
<td>°C</td>
<td>LIM</td>
<td>Tₚₙₙ(d) ≤ Tₖ(d) ≤ Tₚₙₙ(d)</td>
<td>284</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>P(d)</td>
<td>mm</td>
<td>LIM</td>
<td>0 ≤ P(d) &lt; 160</td>
<td>226</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vapor pressure</td>
<td>e(d)</td>
<td>kPa</td>
<td>LIM</td>
<td>0.15 ≤ e(d)eₘₙₙ(Tₚₖ(d)) &lt; 0.96</td>
<td>283</td>
<td>19</td>
<td>(6.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barometric pressure</td>
<td>P_b(d)</td>
<td>kPa</td>
<td>LIM</td>
<td>88 ≤ P_b(d) ≤ 106</td>
<td>173</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

1. SI(d) = 21.77 + 10.63cos(2π·(d−169)/365).
2. ΔSI(d) = 16.10 + 6.80cos(2π·(d−163)/365).
3. Tₚₙₙ(d) = -7.36 + 19.12cos(2π·(d−200)/365).
5. ΔTₚₙₙ(d) = 15.75 + 6.35sin(2π·(d−295)/365).
6. Tₚₙₙ(d) = 0.88 + 9.61cos(2π·(d−199)/365).
7. Tₚₙₙ(d) = 1.77 + 14.30cos(2π·(d−199)/365).
8. Tₚₙₙ(d) = 21.87 + 16.30cos(2π·(d−199)/365).
<table>
<thead>
<tr>
<th>Windspeed</th>
<th>$u_i(d)$</th>
<th>m s$^{-1}$</th>
<th>LIM</th>
<th>$0.45 &lt; u_i(d) &lt; 45.00$</th>
<th>273</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>$</td>
<td>u_i(d) - u_i(d-1)</td>
<td>&lt; 10$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>$u_i(d) = u_i(d-1) = u_i(d-2)$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wind direction</td>
<td>$\theta_j(d)$</td>
<td>°Compass</td>
<td>LIM</td>
<td>$0 \leq \theta_j(d) &lt; 360$</td>
<td>284</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>$\min (</td>
<td>\theta_j(d) - \theta_j(d-1)</td>
<td>, 360 -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>$\theta_j(d) = \theta_j(d-1) = \theta_j(d-2)$</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4
Screening rules for Trevnor daily data

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Symbol</th>
<th>Units</th>
<th>Rule name</th>
<th>Condition for acceptable data</th>
<th>Total</th>
<th>No. excluded (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar irradiance</td>
<td>SI(d)</td>
<td>MJ m(^{-2})</td>
<td>LIM</td>
<td>0 ≤ SI(d) &lt; SI(d)(^{a})</td>
<td>365</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No. excluded (%): 2 (0.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NOC: SI(d) = SI(d-1) = SI(d-2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air temperature</td>
<td>T(_{a})(d)</td>
<td>°C</td>
<td>LIM</td>
<td>T(<em>{a})(d) &lt; T(</em>{a})(d-1) = T(_{a})(d-2)</td>
<td>364</td>
<td>3 (0.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ROC:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NOC:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 10 cm</td>
<td>T(_{v})(d)</td>
<td>°C</td>
<td>LIM</td>
<td>T(<em>{v})(d) &lt; T(</em>{v})(d-1) = T(_{v})(d-2)</td>
<td>365</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 20 cm</td>
<td>T(_{v})(d)</td>
<td>°C</td>
<td>LIM</td>
<td>T(<em>{v})(d) &lt; T(</em>{v})(d-1) = T(_{v})(d-2)</td>
<td>365</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>P(d)</td>
<td>mm</td>
<td>LIM</td>
<td>0 ≤ P(d) &lt; 120</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vapor pressure</td>
<td>e(d)</td>
<td>kPa</td>
<td>LIM</td>
<td>0.15 ≤ e(d) ≤ 0.96</td>
<td>364</td>
<td>12 (3.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barometric pressure</td>
<td>P(_{b})(d)</td>
<td>kPa</td>
<td>LIM</td>
<td>88 ≤ P(_{b})(d) ≤ 106</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\) SI(d) = 21.77 + 10.63\cos(2 \cdot \pi \cdot (d-169)/365), from Ames, IA.

\(^{b}\) ΔSI(d) = 16.10 + 6.80\cos(2 \cdot \pi \cdot (d-169)/365), from Ames, IA.

\(^{c}\) T\(_{a}\)(d) = -5.84 + 18.82\cos(2 \cdot \pi \cdot (d-199)/365),

\(^{d}\) T\(_{v}\)(d) = 27.70 + 12.39\cos(2 \cdot \pi \cdot (d-191)/365).

\(^{e}\) ΔT\(_{v}\)(d) = 12.76 + 6.55\sin(2 \cdot \pi \cdot (d-295)/365).

\(^{f}\) T\(_{v}\)(d) = -2.49 + 15.61\cos(2 \cdot \pi \cdot (d-199)/365), from Ames, IA.

\(^{g}\) T\(_{v}\)(d) = 26.82 + 17.41\cos(2 \cdot \pi \cdot (d-199)/365), from Ames, IA.

\(^{h}\) T\(_{v}\)(d) = 1.77 + 14.30\cos(2 \cdot \pi \cdot (d-199)/365), from Ames, IA.

\(^{i}\) T\(_{v}\)(d) = 21.87 + 16.30\cos(2 \cdot \pi \cdot (d-199)/365), from Ames, IA.
<table>
<thead>
<tr>
<th>Windspeed</th>
<th>$u_i(d)$ m s$^{-1}$</th>
<th>Lim</th>
<th>$0.45 &lt; u_i(d) &lt; 45.00$</th>
<th>357</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Roc</td>
<td>$</td>
<td>u_i(d) - u_i(d-1)</td>
<td>&lt; 10$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NoC</td>
<td>$(u_i(d) = u_i(d-1) = u_i(d-2))$</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Wind Direction</td>
<td>$\theta_i(d)$</td>
<td>C/Comp</td>
<td>$0 \leq \theta_i(d) &lt; 360$</td>
<td>365</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Roc</td>
<td>min$(</td>
<td>\theta_i(d) - \theta_i(d-1)</td>
<td>, 360 -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NoC</td>
<td>$(\theta_i(d) = \theta_i(d-1) = \theta_i(d-2))$</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Variable name</td>
<td>Symbol</td>
<td>Units</td>
<td>Rule name</td>
<td>Condition for acceptable data</td>
<td>Test set</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>------------</td>
<td>-----------</td>
<td>--------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Solar irradiance</td>
<td>SI(h)</td>
<td>MJ m⁻²</td>
<td>LIM</td>
<td>0 ≤ SI(h) ≤ SIₜ网络传播 (d,h)ᵃ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>If SI(h)&gt;0 then ¬(SI(h) = SIₜ网络传播 (h-1) = SIₜ网络传播 (h-2) = SIₜ网络传播 (h-3))</td>
<td></td>
</tr>
<tr>
<td>Air temperature</td>
<td>Tₚ(h)</td>
<td>°C</td>
<td>LIM</td>
<td>Tₚ网络传播 (d)-2.5 ≤ Tₚ(h) ≤ Tₚ网络传播 (d)+2.5ᵇ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(Tₚ(h) = Tₚ网络传播 (h-1) = Tₚ网络传播 (h-2) = Tₚ网络传播 (h-3))</td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 4 cm</td>
<td>Tₐ,h(h)</td>
<td>°C</td>
<td>LIM</td>
<td>Tₐ网络传播 (d)-Tₐ网络传播 (h) ≤ Tₐ网络传播 (d)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(Tₐ网络传播 (h) = Tₐ网络传播 (h-1) = Tₐ网络传播 (h-2) = Tₐ网络传播 (h-3))</td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 10 cm</td>
<td>Tₘ,h(h)</td>
<td>°C</td>
<td>LIM</td>
<td>Tₘ网络传播 (d)-Tₘ网络传播 (h) ≤ Tₘ网络传播 (d)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(Tₘ网络传播 (h) = Tₘ网络传播 (h-1) = Tₘ网络传播 (h-2) = Tₘ网络传播 (h-3))</td>
<td></td>
</tr>
<tr>
<td>Soil temperature at 20 cm</td>
<td>Tₛ,h(h)</td>
<td>°C</td>
<td>LIM</td>
<td>Tₛ网络传播 (d)-Tₛ网络传播 (h) ≤ Tₛ网络传播 (d)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(Tₛ网络传播 (h) = Tₛ网络传播 (h-1) = Tₛ网络传播 (h-2) = Tₛ网络传播 (h-3))</td>
<td></td>
</tr>
<tr>
<td>Precipitation</td>
<td>P(h)</td>
<td>mm</td>
<td>LIM</td>
<td>0 ≤ P(h) ≤ 51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>If P(h)&gt;0 then</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>If P(h)&gt;0 then ¬(P(h)=P(h-1)=P(h-2)=P(h-3))</td>
<td></td>
</tr>
<tr>
<td>Vapor pressure</td>
<td>e(h)</td>
<td>kPa</td>
<td>LIM</td>
<td>0.15 ≤ e(h) ≤ eₜ网络传播 (Tₚ网络传播 (h))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(e(h) = e(h-1) = e(h-2) = e(h-3))</td>
<td></td>
</tr>
<tr>
<td>Barometric pressure</td>
<td>Pₚ(h)</td>
<td>kPa</td>
<td>LIM</td>
<td>88 ≤ Pₚ网络传播 (h) ≤ 106</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>¬(Pₚ网络传播 (h)=Pₚ网络传播 (h-1)=Pₚ网络传播 (h-2)=...=Pₚ网络传播 (h-11))</td>
<td></td>
</tr>
</tbody>
</table>

ᵃ SI网络传播: Extraterrestrial model from Iqbal (1983) using latitude, h, and d.
ᵇ T网络传播 (d) = -7.36 + 19.12cos(2π/(d-200)/365))
ᶜ T网络传播 (d) = 26.15 + 12.14cos(2π/(d-193)/365)).
ᵈ T网络传播 (d) = 26.82 + 17.41cos(2π/(d-199)/365)).
ᵉ T网络传播 (d) = 21.87 + 1630cos(2π/(d-199)/365)).
ᶠ The vapor pressure, e, is calculated from a direct measurement and stored in the CR21's memory. The saturated vapor pressure, eₜ网络传播, is calculated from the Goff-Gratch equation.
| Windspeed | $u_i(h)$ m s$^{-1}$ | LIM | $0.45 < u_i(h) < 45.00$ | 6849 | 140 (2.0) |
| ROC | $|u_i(h)-u_i(h-1)| < 7.5$ | 0 |
| NOC | $-(u_i(h) = u_i(h-1) = u_i(h-2) = u_i(h-3))$ | 91 (1.3) |

| Wind direction | $\theta(h)$ °Compass | LIM | $0 \leq \theta_i(h) < 360$ | 6897 | 0 |
| ROC | $\min (|\theta_i(h)-\theta_i(h-1)|, 360 - |\theta_i(h)-\theta_i(h-1)|) < 150$ | 37 (0.5) |
| NOC | $-(\theta_i(h) = \theta_i(h-1) = \theta_i(h-2) = \theta_i(h-3))$ | 0 |
**Table 6**

Screening rules for Trevnor hourly data

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Symbol</th>
<th>Units</th>
<th>Rule name</th>
<th>Condition for acceptable data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar irradiance</td>
<td>( S_i(h) )</td>
<td>MJ m(^{-2})</td>
<td>LIM</td>
<td>0 ≤ ( S_i(h) &lt; S_{im}(d,h)^a )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>If ( S_i(h) &gt; 0 ) then ( -S_i(h) = S_i(h-1) = S_i(h-2) = S_i(h-3) )</td>
</tr>
<tr>
<td>Air temperature</td>
<td>( T_r(h) )</td>
<td>°C</td>
<td>LIM</td>
<td>( T_{r\max}(d)-2.5 ≤ T_r(h) ≤ T_{r\max}(d)+2.5^b )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>( -T_r(h) = T_r(h-1) = T_r(h-2) = T_r(h-3) )</td>
</tr>
<tr>
<td>Soil temperature at 10 cm</td>
<td>( T_s(h) )</td>
<td>°C</td>
<td>LIM</td>
<td>( T_{s\max}(d) ≤ T_s(h) ≤ T_{s\max}(d)^c )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>( -T_s(h) = T_s(h-1) = T_s(h-2) = T_s(h-3) )</td>
</tr>
<tr>
<td>Soil temperature at 20 cm</td>
<td>( T_s(h) )</td>
<td>°C</td>
<td>LIM</td>
<td>( T_{s\max}(d) ≤ T_s(h) ≤ T_{s\max}(d)^d )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>( -T_s(h) = T_s(h-1) = T_s(h-2) = T_s(h-3) )</td>
</tr>
<tr>
<td>Precipitation</td>
<td>( P(h) )</td>
<td>mm</td>
<td>LIM</td>
<td>0 ≤ ( P(h) &lt; 51 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>If ( P(h) &gt; 0 ) then (</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>If ( P(h) &gt; 0 ) then ( -P(h) = P(h-1) = P(h-2) = P(h-3) )</td>
</tr>
<tr>
<td>Vapor pressure</td>
<td>( e(h) )</td>
<td>kPa</td>
<td>LIM</td>
<td>0.15 ≤ ( e(h)/e_{s}(T_r(h)) &lt; 0.96^e )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>( -e(h) = e(h-1) = e(h-2) = e(h-3) )</td>
</tr>
<tr>
<td>Barometric pressure</td>
<td>( P_b(h) )</td>
<td>kPa</td>
<td>LIM</td>
<td>88 ≤ ( P_b(h) ≤ 106 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ROC</td>
<td>(</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NOC</td>
<td>( -P_b(h) = P_b(h-1) = P_b(h-2) = \ldots = P_b(h-11) )</td>
</tr>
</tbody>
</table>

---

1. \( S_{im} \): Extraterrestrial model from Iqbal (1983) using latitude, h, and d.
2. \( T_{r\max}(d) = -5.84 + 18.82\cos(2 \cdot \pi \cdot (d-199)/365) \).
3. \( T_{s\max}(d) = 27.70 + 12.39\cos(2 \cdot \pi \cdot (d-191)/365) \).
4. \( T_{s\max}(d) = -2.49 + 15.61\cos(2 \cdot \pi \cdot (d-199)/365) \), from Ames, IA.
5. \( T_{s\max}(d) = 26.82 + 17.41\cos(2 \cdot \pi \cdot (d-199)/365) \), from Ames, IA.
6. \( T_{s\max}(d) = 21.87 + 16.30\cos(2 \cdot \pi \cdot (d-199)/365) \), from Ames, IA.
7. \( e \) and \( e_s \) are calculated from direct measurements and stored in the CR21's memory.
<table>
<thead>
<tr>
<th>Windspeed $u_i(h)$ m s$^{-1}$</th>
<th>LIM $0.45 &lt; u_i(h) &lt; 45.00$</th>
<th>8740</th>
<th>52</th>
<th>(0.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC $</td>
<td>u_i(h)-u_i(h-1)</td>
<td>&lt; 7.5$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>NOC $u_i(h) = u_i(h-1) = u_i(h-2) = u_i(h-3)$</td>
<td>24</td>
<td>(0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind Direction $\theta_i(h)$</td>
<td>LIM $0 \leq \theta_i(h) &lt; 360$</td>
<td>8760</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ROC $\min (</td>
<td>\theta_i(h)-\theta_i(h-1)</td>
<td>, 360 -</td>
<td>\theta_i(h)-\theta_i(h-1)</td>
<td>) &lt; 150$</td>
</tr>
<tr>
<td>NOC $\theta_i(h) = \theta_i(h-1) = \theta_i(h-2) = \theta_i(h-3)$</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. (continued)
Fig. 1. (a) The LIM rule for daily air temperature data at Walnut Creek. (b) The ROC rule for daily air temperature data at Walnut Creek.
Fig. 2. The LIM rule for hourly solar irradiance data at Walnut Creek.
A NOTE ON USING THE SAMSON DATABASE FOR ESTIMATING LINKE TURBIDITY FACTOR

A note prepared in the American Meteorological Society format

D.W. Meek

ABSTRACT

For the 30 year period from 1961-1990, the Solar and Meteorological Observation Network (SAMSON) database provides modeled daily broadband aerosol optical depth data and precipitable water values for hundreds of locations throughout the United States of America. These data can be used to model the Linke turbidity factor. There has been, however, no confirmation assessment reported with independent long-term data for the Linke turbidity factor and only limited assessment of the SAMSON aerosol optical depth data. The purpose of this note is to report results of a comparison of published long-term monthly mean values of Linke turbidity factor recorded for Tucson, AZ to corresponding ones estimated from the closest SAMSON site. Monthly mean Linke turbidity factor values were obtained from tabulated values in a Zumber and Sellers' publication. Daily Linke turbidity factor estimates for the nearby SAMSON site, the Tucson Airport National Weather Service Office, were based on noontime values of broadband aerosol optical depth, precipitable water, and air pressure. Monthly mean SAMSON values were then computed from the daily based Linke turbidity factor values. There were 269 points in common from January 1961 through June, 1983. Methods of comparison included graphs, descriptive statistics, and measurement error regression analysis. The SAMSON based Linke turbidity factor estimates were not different from the Zumber and Sellers' published observations (P>0.1); the result implies that the SAMSON...
aerosol optical depth data are also viable. The result may be site specific, so similar assessments for other locations would be desirable.

1. Introduction

Assessment of atmospheric turbidity and aerosols is important not only in the study of radiative transfer but also in the study of many related atmospheric environmental issues (see e.g., d'Almeida et al. 1991). Some of the related areas of general interest include climate change, air pollution, atmospheric chemistry, and atmospheric optics. Personal interests include the need to use broadband aerosol optical depth and precipitable water estimates in several different broadband solar shortwave irradiance models (e.g., see Bird and Hulstrom 1981).

Climatic norms and extremes are generally based on a 30-year period of record (WMO 1967). To bring solar radiation data up to this standard (and in cooperation with the U.S. National Climatic Data Center [NCDC]), the U.S. National Renewable Energy Laboratory (NREL, formerly SERI [Solar Energy Research Institute]) developed the National Solar Radiation Data Base which contains a quality controlled 30-year record; it is also known as SAMSON (for Solar and Meteorological Surface Observation Network). The data record is serially complete and includes global and component solar shortwave irradiance data along with broadband aerosol optical depth ($\sigma_a$) and precipitable water ($w$) on an hourly basis for 56 primary sites and 183 secondary sites (NREL 1992, 1995).

In the SAMSON database the $w$ values are interpolated in time and space from the regular NWS sounding data. The $\sigma_a$ values are generally modeled but with some exceptions. At the primary stations the solar radiation components, including $\sigma_a$ calculated from a direct normal measure, are based, at least partially, on radiometer measurements for selected dates throughout a several year model development period. A narrower selection of primary sites and data was used to make an assessment of the model. An evaluation of the modeled $\sigma_a$ against a long term set of
independent measurements, a confirmation set, was not reported in NREL (1995). Moreover, no short or long-term SAMSON-based assessment of a closely related measurement, Linke turbidity factor ($T_n$), has been reported. While a long-term independent assessment of SAMSON $\sigma_n$ is not readily possible, there is a long-term independent record at Tucson, AZ, for monthly $T_n$. The purpose of this note is to present a confirmation analysis for Tucson that compares long-term monthly mean $T_n$ estimates from SAMSON data with corresponding observation based values of $T_n$ tabulated by Zumber and Sellers (1985).

2. Data and Methodology

Since long-term records of turbidity measurements are spatially and temporally sparse, development of a database founded totally on measurements was not possible. The production of the SAMSON database made use of the clear sky algorithms of Bird and Hulstrom (1981), with the exception of aerosol transmittance. Aerosol transmittance used a broadband form of Beer's Law. For SAMSON, Maxwell and Myers at NREL developed a daily-based broadband turbidity surrogate, $\sigma_n$ (Chapter 6, NREL 1995), based on Beer's Law applied to direct broadband atmospheric transmission adjusted for all other molecular transmission factors in the model (ozone absorption, Rayleigh scattering, uniformly mixed gas absorption, and water vapor absorption). Since Beer's Law is strictly valid only for spectral losses from the beam, arguments and comparisons to spectral models are provided in the NREL documentation.

Zumber and Sellers (1985) reported 27 years of statistics on monthly Linke Turbidity Factor, $T_n$, measured at the University of Arizona (UA). The observation location was on top of the UA Physics and Atmospheric Science Building (coordinates are 32.23° N, 110.95° W, and 760 m elevation). The period of record was from June, 1956 through June, 1983, so a monthly comparison of averages was possible with data from the Tucson SAMSON primary site, Tucson airport National Weather Service (NWS) Office
(WBAN 23160, coordinates are 32.12° N, 110.92° W, and 779 m elevation).
The period of record for the entire SAMSON database was from January 1961
through December 1990 but the SAMSON database development was generally done on
clear days during the period of record for the direct normal radiometer
measurements, which was less than three years (from January 1, 1978 through
September 2, 1980). The Linke turbidity factor values tabulated in Zymber
and Sellers' (1985) were not adjusted for water vapor although their
analysis and discussion addressed its effects. Consequently the
corresponding estimates developed from the SAMSON database need to include
water vapor as well as aerosols. The SAMSON database includes hourly
estimates for \( c_a \) (dimensionless), \( w \) (mm), and \( p_a \) (atmospheric pressure,
mb). Equating the argument in the Linke transmittance function to the
corresponding arguments in the product of the aerosol and precipitable
water direct beam transmittances for the Bird model (NREL 1995) then
solving for the Linke turbidity factor \( T' \), yields the following estimator
based on data available in the SAMSON database:
\[
T' = (c_a - \ln(T_v(w))) (mP(m) \log(e)) ,
\]
where \( m \) is optical air mass, \( T_v(w) \) is water vapor transmittance, and \( P(m) \) is
the Linke turbidity coefficient. For the actual calculations \( m \) was estimated
with the model of Kasten and Young (1989) and adjusted for elevation with \( p_a \).
Daily noon time values were used to minimize errors due to \( m \). For water
vapor transmittance \( T_v(w) \), the modified model of Bird NREL (1995) was used; it
is as follows:
\[
T_v(w) = 1 - 1.688X_w[(1 + 54.6X_w)^{2.637} + 4.042X_w]^{-1},
\]
where \( X_w = w m \) with \( w \) in cm. Data in Table 4 in the Radiation Commission of
the IAM (1958) were used to develop an interpolating curve for the Linke
turbidity coefficient, \( P(m) \). The resulting \( P(m) \) function developed to model
the Linke turbidity factor \( T_v \) with the SAMSON data is given by Eq. (1).
\[
P(m) = 2.1250 + 21.629(m + 2.1056\times10^{-3})^{-1}
\]
All parameters estimated were \( P<0.0001 \) and \( R^2=0.999 \).
Graphical and statistical analyses were used to examine the extent of deviation from a one-to-one relationship. The results of Berg (1992) which includes univariate descriptive statistics in the form of box plots for the observations, the predictions, and their difference (residuals from the one to one line) were used as a means to visually appraise the relationship. Further analyses proceeded with residual time plots, model performance statistics suggested by Fox (1981), the Pearson correlation coefficient, and measurement error regression lines using the method of least normal squares as well as the methods of Draper (1991) and Kerrich (1966). Measurement error regression analysis reduces the ordinary least squares estimation bias in the true, i.e., "structural" relationship (e.g., Fuller 1987).

3. Results and Discussion

The two Tucson data sets had 269 common points (r=0.46, P<0.0001). Performance analyses were redone with various subsets as follows: excluding only the Tucson SAMSON σₜ confirmation period, excluding the top 5% of the modeled set, excluding the top 5% of absolute value differences, including only the SAMSON σₜ development period, and including data only in the lower quartile of monthly mean w. In almost all cases the modeled data slightly under-estimated Zymber and Sellers' measures by about 1% based on the mean predicted value. A slight but insignificant systematic error (P=0.11) was evident based on the relationship between the UA observations of monthly mean Tₐ and the corresponding SAMSON based T (Fig. 1). The model performance statistics for the complete set of data (all 269 data pairs) were as follows: mean observed Tₐ = 2.20; mean predicted Tₐ = 2.17; the mean difference, i.e. mean bias error-MBE, ΔTₐ = 0.03±0.02 (= 1% of the predicted mean); the root mean square error, RSME = 0.34 (= 16% of the predicted mean); and the mean absolute error, MAE = 0.27 (= 12% of the predicted mean).

The residual time plots show the slight bias from different perspectives and highlight other phenomena (Figs. 2 and 3). In all the figures, the open circles are the minimum of all the UA observed values for
each month of the period in common while the open diamonds are the corresponding maximum. Most of these extreme UA observations tended to be associated with the largest residuals. Ten of the 12 highest values occurred after July, 1982 (Fig. 2). The El Chichón Volcano erupted in the spring of that year and resulted in a long duration of unusually high aerosol concentration in the upper atmosphere over the Northern Hemisphere (e.g., see Dutton and DeLuisi 1983). A further discovery is shown in Fig. 3, an undesirable seasonal pattern in the residuals. The importance of this problem to a given application should be assessed by the researcher.

While there are possible systematic and random errors in both values, the available source information on each variable was not detailed enough to exactly assess the measurement error in each of the variables. Measurement error lines were estimated using three different methods based on different assumptions. The Draper regression line is shown as the solid line segment in Fig. 1; Draper recommends this method in the absence of information on the measurement error. The coefficients of the Draper line are given in Eq. (2).

\[ T_{n,\text{observed}} = -0.05 + 1.04T_{n,\text{predicted}} \tag{2} \]

The least squares line is shown as the dashed line in Fig. 1; the corresponding coefficients are given in Eq. (3).

\[ T_{n,\text{observed}} = -0.14 + 1.07T_{n,\text{predicted}} \tag{3} \]

Notice the intercepts in Eqs. [2] and [3] are extrapolated out of the data range and are for practical purposes insignificant. The Kerrich line (not shown on Fig. 1), which forces the line through origin, is given by Eq. [4];

\[ T_{n,\text{observed}} = 1.01T_{n,\text{predicted}} \tag{4} \]

the estimate is about the same as the one from the mean difference. No data were excluded from the regression lines. Hence based on the regression slopes, the predicted data were systematically less than the observations only by 1% or more, probably not much more, because the slope in Eq. (3) may be over estimated (e.g., see Carroll and Ruppert 1996). The magnitude of the under-estimate is insignificant (P>0.1) based on either the Kerrich slope or MBE.
An assessment of the portion of mean predicted $T_a$ factor contributed by the $w$ component indicated that it accounts for 60% of the value; hence, the $\sigma$ component accounts only for 40% of it. Data for $w$ can be considered valid for a large area (NREL 1995). Assuming the SAMSON $w$ data were reasonably the same for the UA site, then by inference the SAMSON $\sigma$ data may be considered as confirmed.

The relationship for Tucson may be site specific, representing only one arid location at a medium elevation. Different data to reproduce similar assessments at other locations are needed. Long-term normal or extreme turbidity estimated from SAMSON data for a different time base or for other locations with different climates and/or altitudes may have a more significant bias.

4. Conclusions

The monthly based SAMSON $T_a$ for Tucson, AZ under-estimated the corresponding independent measurements from Zymbur and Sellers' (1985) by only about 1%, an amount not significantly different from zero ($P<0.1$), indicating very reasonable agreement. While the exact magnitude of the relationship was uncertain because there are unreported measurement errors when using either $T_a$ as the predictor, it was acceptably small by any measure. Although not directly assessed, the SAMSON $\sigma$ by inference should be unbiased. Whether or not the result applies to turbidity parameter estimates from climatic extremes or other time scales and/or locations is unknown. Comparison with observations recorded at other locations would be desirable.

Acknowledgments. This work was supported by the USDA-ARS National Soil Tilth Laboratory, Dr. J.L. Hatfield, Director. Prof. D.N. Yarger, ISU, Ames, IA, provided advice, encouragement, and comments on the manuscript. Dr. E. Dutton, USDC-NOAA-ERL, Boulder, CO, Dr. G. Johnson, USDA-ARS, Boise, ID, Prof. W.D. Sellers, UA, Tucson, AZ, provided comments on earlier versions of the manuscript. Dr. E.L. Maxwell, USDE-NREL, Golden, CO, also provided
comments on an earlier version of the manuscript and information on the SAMSON database. Mrs. J. Meek graciously edited the work.

REFERENCES


Kerrich, J.E., 1966: Fitting the line $Y=ax$ when errors of observation are present in both variables. *Am. Statist.*, 20, 24.


FIG. 1. Comparison of monthly Linke turbidity factor from Zymber and Sellers' (1985) long-term research at the University of Arizona and corresponding estimates from the Tucson NWS SAMSON site (WBAN Station 23160). Monthly data in common were for the period 1961-1983.
FIG. 2. Differences of observed and predicted values for all Fig. 1. points plotted against year. A O represents the residual associated with the lowest observed value for each month, a O the highest. Some of the large differences are points associated with volcanic aerosol emission events.
FIG. 3. Differences of observed and predicted values for all Fig. 1. points plotted against month. A ◊ represents the residual associated with the lowest observed value for each month, a ◇ the highest. Some of the large differences are points associated with volcanic aerosol emission events. A seasonal trend is apparent with the spring values above the zero line and the late fall values below the zero line.
ESTIMATION OF MAXIMUM POSSIBLE DAILY GLOBAL SHORTWAVE SOLAR RADIATION

A paper accepted with revision by Agricultural and Forest Meteorology

D.W. Meek

Abstract

The estimation of maximum possible daily solar radiation is important in many applied sciences. This study develops and evaluates climatic extreme based modifications of two single-atmospheric-layer broadband shortwave irradiance models for the purpose of estimating a dynamic upper boundary for global solar radiation at any given location. Climatic component models were developed for five rural locations in the central United States: Ames, IA, Bismarck, ND, Columbia, MO, Dodge City, KS, and Wooster, OH. Each site had long-term (30 or 31 years) records of daily global solar radiation data available. Aerosol optical depth, precipitable water, and surface albedo were the input variables. Data for the first two inputs were obtained from the SAMSON database (Solar and Meteorological Observation Network). Albedo interpolating curves were estimated from the predecessor of SAMSON. For each site, precipitable water and aerosol optical depth daily data were used to develop annual trends in the climatic lows and normals for each variable. The normals were based on median daily values. Nonlinear generalized least squares regression analyses were used to develop the interpolating curves. To evaluate the broadband solar shortwave radiation models, the maximum daily solar radiation values for a given day from the entire period of record for each site and day in the year were selected. In either broadband model, the use of the climatic normals in each input variable either interpolated or under-estimated the selected radiometer data. The use of the climatic lows, however, did yield

\[1\] Contribution of the Agricultural Research Service, U.S. Department of Agriculture. Submitted as part of the dissertation requirement of Iowa State University.

\[2\] Mathematician, United States Department of Agriculture, Agricultural Research Service-Midwest Area, National Soil Tilth Laboratory, 2150 Pammel Drive, Ames, IA 50011-4420, USA.
a reasonable upper boundary for the selected data. The result may be partially due to the fact that the climatic minima curves for each turbidity variable were generally significantly different throughout the year from the climatic normals (P < 0.05). The models were most sensitive to aerosol optical depth. Although it is more sensitive to input variation and it is somewhat less conservative, the simpler of the two broadband models is adequate for most applications. While results are site specific, the methodology is general.

1. Introduction

Estimation of maximum possible daily clear sky solar radiation at the earth's surface is needed for agronomic modeling and managerial purposes. Solar radiation is increasingly being measured in agricultural weather data networks (e.g., see Meyer and Hubbard, 1992 or Snyder and Pruitt, 1992). Meek and Hatfield (1994) propose screening daily global solar shortwave irradiance data, SI(d), with a dynamic upper boundary called a "clear day curve", SI_c(d), which should equal the maximum possible daily values for the measurement. So for each daily datum, SI(d) ≤ SI_c(d). Other needs for SI_c(d) include models used for estimating the longwave portion of surface net radiation and models for estimating cloudiness or attenuation which contain a cloudiness factor f (e.g., see Brutsaert, 1982; Davies and Idso, 1979; Dong et al., 1992; Howell et al., 1984; Jensen et al., 1990; or Shuttleworth, 1993). One parameterization for f is f = a(SI(d)/SI_c(d)) + b where a and b are longwave radiation coefficients for clear skies (with a+b=1). Furthermore other climatic, environmental, solar energy, and engineering applications need such an estimate (e.g., see Iqbal, 1983).

Clear day curves can be crudely inferred from charts in a solar atlas (e.g., see USDOE, 1978 or Baker and Klink, 1975). They are sometimes developed from site specific long-term radiometer data, but such data are not widely available at many sites like those within the Cooperative Station Network. Moreover, Meyer and Hubbard (1992) or Snyder and Pruitt (1992) list many automated stations that have more or less recently come
into operation and so have limited historical records. When long-term data are available, sinusoidal regression curves that interpolate these data have been used for estimating clear day curves (e.g., see Meek and Hatfield, 1994 or Heermann et al., 1985). Ideally such a curve should not interpolate the data but bound them in the mathematical sense of a least upper bound for every day in the year. An alternative method would be to use radiative transfer models set for climatic extremes of favorable transmission conditions. Unfortunately, for spectral models, geographically representative data are not readily available. Conceptually, as a more practical alternative, clear sky adaptations of single-atmospheric-layer broadband radiative transfer routines (generally, .3 - 3.0 μm) similar to those of Atwater and Ball (1976) or Iqbal (1983) should do but their use has also been limited because they require input data that have been scarce or unavailable and had to be estimated in various, often indirect, ways (e.g., see Suckling and Hay, 1976; Meyers and Dale, 1983; or Tracy et al., 1983).

Now, however, with the availability of the Solar and Meteorological Observation Network database (SAMSON) the situation has changed. To bring solar radiation data up to the 30-year WMO (1967) standard (and in cooperation with the U.S. National Climatic Data Center [NCDC]), the U.S. National Renewable Energy Laboratory (NREL, formerly SERI [Solar Energy Research Institute]) developed the National Solar Radiation Data Base (NRSDB) which is more commonly called the SAMSON database. It contains a quality controlled 30-year record. The record includes total and component shortwave irradiance data along with many other meteorological variables, all on an hourly basis, for 56 primary sites and 183 secondary sites (NREL, 1992; 1995). The SAMSON database radiation components at the primary stations are based, at least partially, on measurements, while corresponding data for the secondary stations are modeled. The production of the SAMSON database made use of the clear sky algorithms of Bird and
Hulstrom (1981), with the exception of aerosol transmittance. Aerosol transmittance used a broadband form of Beer’s Law.

The SAMSON database can be employed in a variety of useful ways including the previously mentioned solar radiation and surface energy balance applications used by a variety of research, public, and private sector organizations. Meyer and Hubbard (1992) reviewed the data needs and development of nonfederal weather station networks; Snyder and Pruitt (1992) reviewed evapotranspiration data needs for irrigation and drainage management; and some recent multi agency hydrological projects required short-term real-time meteorological data (e.g., see USDA, 1994 or Kustas and Goodrich, 1994).

My goal is to estimate maximum possible daily global solar shortwave irradiance at an arbitrary location with a broadband model. Ideally such a model estimates the upper boundary for radiometer data from an instrument set on a horizontal surface and in a very clear, cloud-free, and dry atmosphere. The hypothesis is that using the trend in climatic low for each climatic input component with a chosen broadband model will do a better job of bounding selected clear day radiometer observations than using the corresponding normal trends in the same chosen broadband model. To accomplish my goal, I adapted two hourly-based clear-sky single-atmospheric-layer broadband radiative transfer models for use with annual trend climatic component input models for the turbidity inputs, precipitable water and aerosol optical depth.

2. Data

A summary of data used in this study is presented in Table 1. Aerosol optical depth, precipitable water, and albedo data were needed to develop the annual trend models that were to be used in the selected broadband global shortwave solar radiation models. Five small-town, more or less rural, sites were chosen where radiometer data were directly available and the climatic data for developing the annual trend component
models were either directly available or could be reasonably interpolated from the closest neighbors. The sites chosen were Ames, IA, Bismarck, ND, Columbia, MO, Dodge City, KS, and Wooster, OH. Each of the considered climatic inputs may reasonably, in given circumstances, be interpolated over large areas: for aerosol optical depth see e.g., Flowers et al. (1969), Iqbal (1983), or NREL (1995); for precipitable water see e.g., NREL (1995); and for albedo see e.g., Hummel and Reck (1979) or Kung et al. (1964). Moreover, based on studies as well as theory, very clear day radiation should be uniform over wide areas, especially longitudinally in areas of low relief (Bland and Clayton, 1994). The daily Ames and Wooster global solar radiation data with limited additional meteorological data were obtained from the agricultural experiment stations in Iowa and Ohio. The SAMSON database and its predecessor, called the SOLMET database, were the source for the other data. The three chosen SAMSON sites, Bismarck, Columbia, Dodge City, are all well-known and documented in the literature (e.g., see NREL, 1992 and 1995; Baker and Klink, 1975; or Iqbal, 1983). In addition these sites were selected because they are SAMSON primary stations which are located in mainly rural areas. Omaha, Madison, and Columbia were selected as the nearest primary stations for interpolating ancillary SAMSON data for use in the Ames broadband model. Akron/Canton was selected as a secondary SAMSON station that is nearest Wooster; so near, that no interpolation schemes were needed.

2.1. Aerosol optical depth (\( \sigma_a \))

A database for turbidity entirely founded on measurements was not possible because long-term records of turbidity measurements are spatially and temporally sparse. Hence, for the SAMSON database, Maxwell and Myers at the NREL developed a broadband turbidity surrogate (broadband aerosol optical depth - \( \sigma_a \)) based on Beer's Law applied to direct broadband atmospheric transmission (Chapter 6, NREL, 1995). Arguments and comparisons to spectral models are provided in the NREL documentation.
because Beer's Law is strictly valid only for spectral losses from the beam.

Since neither Ames nor Wooster were in the SAMSON database, to obtain \( \sigma_a \) for Ames, IA, data from the SAMSON Primary Stations at Omaha, NE (WBAN 94918), Columbia, MO (WBAN 03945), and Madison, WI (WBAN 14837) were averaged with inverse distance weights. The respective weights were 0.449, 0.279, and 0.272. Since all stations are within a 100 m (<0.01 standard atmosphere) elevation difference from Ames, no elevation adjustments were made. While Des Moines, IA (the nearest station in the SAMSON database) is a nearby secondary station, the primary sites were deliberately selected. The criteria were: (1) the \( \sigma_a \) data were at least partially based on solar radiation measurements and (2) the resulting Ames \( \sigma_a \) model can then be compared to the model NREL developed for Des Moines (Appendix B, Table B1, NREL, 1995). While data for Wooster could have been interpolated from the nearest primary stations, the direct use of data from the closest station was of obvious practical interest. Hence to test the direct use of data from a nearby secondary station, \( \sigma_a \) data from Akron/Canton, OH (WBAN 14895) were used for \( \sigma_a \) data at Wooster. Bismarck, Columbia, and Dodge City daily \( \sigma_a \) were directly used. In all cases, the daily data values were the same as the noon time hourly values.

2.2. Total precipitable water (w)

Surface dew point or humidity has been used to estimate \( w \) in broadband models (e.g., Meyers and Dale, 1983). Unfortunately, there are problems with this approach. Dew point data or sounding data are often unavailable at some Class A or lower quality sites, especially some of those in the Cooperative State Network. Moreover Reber and Swope (1972) found that \( w \) is often poorly related to surface humidity. Without a sounding or other test the validity of the \( w \) estimate from the surface observation is unknown. Moreover \( w \) was conveniently included in the meteorological data for the SAMSON database and was, therefore, directly
used. Average daily w values over the 24 hour period were computed from the hourly values for each of the chosen SAMSON sites. Estimation of the Ames daily w value followed the interpolating procedure used in estimating the daily Ames $\sigma_a$ value.

2.3. Surface albedo ($\alpha$)

Albedo data are not widely available. Interpolating values can present problems since local surface conditions could vary considerably but in the absence of local data, there was no measurement-based alternative. Using data from SAMSON’s predecessor, SOLMET, and other sources, Iqbal (1983) tabulates monthly means for numerous locations throughout the United States and Canada. No extremes or variations were given. Included were values for the three primary stations used to obtain the Ames $\sigma_a$ values. The respective monthly $\alpha$ values at the primary sites and previously calculated weights were used to estimate monthly $\alpha$ values for Ames. Data for Wooster were interpolated from Columbia and Madison in the United States and Toronto in Canada. Respective inverse distance-based weights were 0.122, 0.218, and 0.660. Data for Bismarck, Columbia, and Dodge City were used directly.

2.4. Global shortwave irradiance (SI)

The daily SI data for Bismarck, Columbia, and Dodge City were chosen because network history, operations, instrumentation, changes, and calibrations associated with the data were well-documented (Chapter 2, NREL, 1995). The period of record was from 1961-1990. Corresponding data for Ames, IA recorded from 1960 through 1990 were obtained from the Iowa State University (ISU) Experiment Station. Similar data for Wooster, OH recorded from 1962 through 1992 were obtained from the nearby Ohio State University Experiment Station (OARDC). At both the ISU and OARDC sites an
The Wooster radiometer was located on the OARDC farm. Additional information, especially for Wooster, was sparse. The Ames radiometer was located on the ISU Campus but was moved three times during the period of record because the responsible department was relocated. Complete instrumentation and calibration history has not been formally summarized but a brief assessment of the data quality was reported in Baker and Klink (1975). Up until the early 1980s data were recorded on strip charts and daily data values were obtained via planimetry. The methodology could have introduced a 5% or more error, possibly tending toward systematic overestimation of the daily values.

3. Methodology

Annual trend regression models for each site (i.e., depending only on day in the year) were developed to provide aerosol optical depth ($c_\alpha(d)$), precipitable water ($w(d)$), and albedo ($\alpha(d)$) inputs for two different broadband solar radiation routines. To avoid problems with unknown underlying distributions in the data and inherent heteroscedasticity, generalized least-squares regression (GLS a.k.a. "weighted") analyses were used for the $c_\alpha(d)$ depth and $w(d)$. Heteroscedasticity is most commonly characterized by nonconstant standard error of the regression estimate but there are many other important statistical issues involved with this property (e.g., see Carroll and Rupert, 1988). Ordinary least squares regressions (OLS) were used with the $\alpha(d)$ data. In each regression analysis, several diagnostics, including residual analyses, were examined.

---

3 Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.
3.1. Broadband model selection

The first solar radiation model, hereafter called "model 1", was based on Bird and Hulstrom's model (1981). It includes some of the modifications by NRSDB (NREL, 1995) and some proposed by Iqbal in his broadband model 3 (Iqbal 1983). Iqbal (1983) and Bird and Hulstrom (1981) argue that this model compares well with various spectral models. The basic equation for total clear sky shortwave irradiance (direct + diffuse) was

\[ S_{I_c} = I_c \cos(Z) T_R T_a T_d T_s + 0.79 T_n (0.5(1-T_d) + F_c (1-T_n)) (1-m+0.62^{-1}) (1-r_d)^{-1} \]

where the transmission equations and associated parameters are

\[ m = (\cos(Z) + 0.50572(96.07995-Z))^{-1} \]
\[ m_a = mP/P_0 \]
\[ T_R = \exp(-0.0903 m^2(1+5.8-0.001^2)) \]
\[ U_0 = 1 - 0.1611 U_a (1 + 139.48 U_a)^{-0.3035} - 0.002715 U_o (1 + 0.044 U_o - 0.0003 U_o^2)^{-1} \]
\[ U_o = w m_r \]
\[ T_a = 1 - 2.4959 U_a ((1 + 79.034 U_a)^{0.6928} + 6.385 U_a)^{-1} \]
\[ T_s = \exp(-\sigma m) \]
\[ T_n = 1 - (1-w_o) (1-m+0.62^{-1}) (1-T_d) \]
\[ T_{n_s} = T_s / T_n \]
\[ r_s = 0.0685 + (1-F_c) (1-T_{n_s}) \]

The air mass equation was from Kasten and Young (1989). The optical path for ozone and the dimensionless water vapor absorption were from Iqbal (1983). While Iqbal (1983) had seasonally varying norms for total ozone thickness, \( U_o \), a constant value of 0.3 cm was used because the variation was small compared to other factors. As in Bird and Hulstrom (1981) and Iqbal (1983), the single scattering albedo, \( w_o \), was set to 0.91. The ratio of forward to total aerosol scattering, \( F_c \), was set to 0.82 (rural setting). Therefore model 1 has direct beam irradiance that includes the effects of Rayleigh scattering and absorption by ozone, water vapor, permanent gases, and aerosols. It also has diffuse irradiance due to single Rayleigh scattering, aerosol absorption and scattering, and a
correction factor for multiple surface reflections of single scattered radiation. Model 1 was chosen because it contains physically based terms comparable to a reasonably comprehensive spectral model.

The second solar radiation model, hereafter called "model 2", was based on Bird and Hulstrom's (1981) modifications to the model of Atwater and Brown (1974). It has been employed in several studies by Atwater and Ball (1976; 1978a, b; and 1981) and reportedly is valid for very clear atmospheric conditions (Bird and Hulstrom, 1981). The basic equation for total clear sky shortwave irradiance (direct + diffuse) was

\[ S_{I_c} = I_0 \cos(Z) (T_H - a_m) T_e (1 - r_a)^{-1} \]

where the transmission equations and associated parameters are

\[ m = 35 (1224 \cos^2(Z) + 1)^{-0.4} \]
\[ m_r = \frac{m}{P_0} \]
\[ T_H = 1.021 - 0.0824 \{ (949 \times 10^{-5} P + 0.051)^{0.5} \} \]
\[ a_m = 0.077 (U^m)^{0.3} \]
\[ T_e = \exp(-\sigma_a m) \]

Model 2 uses Rodgers' air mass model (1967) but with absolute air mass instead of relative air mass in the aerosol transmission component so it was consistent with the corresponding term in model 1. Hence model 2, as parameterized, includes direct and diffuse irradiance. It accounts for effects due to aerosols and due to molecular absorption effects except water vapor. Water vapor absorptance is included separately. Also it has a factor for multiple surface reflections of single scattered radiation. Model 2 was chosen because it has few specific parameterizations and so provides a good contrast to model 1.

Solar constant, equations of time, and solar hour angle were developed according to Iqbal (1983) and were the same for both models. Hour angle was centered in the one-hour time intervals including sunrise and sunset hours. For the purpose of this paper, adjusting the photoperiod and centering the hour angle in the sunrise and sunset photoperiod would be of little gain because the resulting bias in both daily solar radiation...
model estimates were both median bias error < 0.01 MJ m\(^{-2}\) based on Ames normal conditions. The maximum bias was |\<0.25| MJ m\(^{-2}\). Maximum bias values were associated with extremes in the yearly variation of the equation of time. Daily totals were the sum of hourly values throughout the period of sunshine. Aside from \(c_a\), \(w\), and \(\alpha\), the remaining inputs depended on the geographic coordinates. Standard atmospheric pressure and absolute air mass were adjusted for elevation via the hypsometric equation.

3.2. Aerosol optical depth analyses

A comparison of monthly Linke turbidity factor (\(T_n\)) constructed from the \(c_a\) and \(w\) values for the Tucson SAMSON site (WBAN 23160) with the long-term \(T_n\) observations (27 years) of Zymber and Sellers (1985) at the nearby University of Arizona revealed no significant systematic differences (unpublished research). Hence the daily \(c_a\) and \(w\) data for the chosen sites in this paper were directly used without any adjustment. Median values of aerosol optical depth were determined over all years by each day within the year and then used to represent the normal value, \(c_{\text{a,normal}}(d)\). Similarly, corresponding minimum values were used to represent the climatic lower boundary, \(c_{\text{a,lower}}(d)\). Fourier series and other suitable nonlinear candidate models for GLS regressions were used to develop annual trends. For each radiometer site selected regression models were used to estimate the normal and the minimal predictions along with their 95% confidence limits.

3.3. Precipitable water analyses

Precipitable water data extraction for each site paralleled that for aerosol optical depth. Similarly medians represent normal precipitable water, \(w_{\text{normal}}(d)\), and minima represent the climatic lower boundary, \(w_{\text{lower}}(d)\). Again, GLS regression methodology was employed. Candidate models included Gaussian forms (bell shaped) as well as Fourier series.
3.4. Albedo analyses

Since only mean monthly values were available and statistical
uncertainty was not reported, ordinary least squares (OLS) methodology was
used for the albedo regression analyses. The middle day of the month was
used for the time scale. To track the shape in the albedo scatter plots
for each site (which have high values in the winter and low values in the
summer), models considered included Fourier series and splined polynomials.

3.5. Broadband model analyses

To test the potential clear day model predictions of the climatic
upper boundary at each site, maximum values determined over all years by
each day in the year were used \([S_{\text{max}}(d)]\); hereafter, these data are called
the ‘selected observations’. A summary of what the selection procedure
captured will precede an assessment of the modeling. For each site the
selected data were subtracted from the model 1 and 2 predictions with three
variations on \(\sigma_i\) and \(w\): (1) the “normal” curves for each were used (i.e.,
\(\sigma_{\text{norm}}(d)\) and \(w_{\text{norm}}(d)\)); (2) the climatic “minimum” curves for each were used
(i.e., \(\sigma_{\text{min}}(d)\) and \(w_{\text{min}}(d)\)); and (3) the 95% lower confidence limit for each
of the minimum curves were used (i.e., \(\sigma_{\text{all}}(d)\) and \(w_{\text{all}}(d)\)). The latter
case was designated “LL1”. For each site univariate statistics for each of
the differences were calculated. In addition, a simple error analysis for
each of the inputs, \(\alpha\) included, was conducted to assess the systematic
variation in broadband predictions due to systematic differences in each of
the input component estimates. For the Ames normal case, errors for each
were set: \(\Delta\sigma_i = 0.01; \Delta w = 1 \text{ mm}; \) and \(\Delta\alpha = 0.1\). Again univariate statistics
for each variable were calculated.

Finally, for each model with a reselection from the selected data for
the very clearest day data from all the three rural primary SAMSON sites,
a further, more robust, assessment analysis was done. For a “reselected”
set and predictions for each model from the climatic “minimum”
parameterization, the plot of Berg (1992) along with time residual plots
and analyses, the Pearson correlation coefficient, and a combination of other statistics suggested by Fox (1981) and Willmott (1982) were used to assess the bounding of the observations by each model. The other statistics were both variable means, mean bias error (bias), mean absolute error (MAE), root mean square error (RMSE), variance of the bias ($s^2$), and OLS linear regression analysis. There was, however, a major difference for Wilmott's statistics; observations were treated as the dependent variable. This choice was made because the observations contained random errors. In contrast, model predictions were independent of random errors because inputs were all based only on time trend regressions. Model predictions only have systematic errors. These analyses are generally intended to confirm that models adequately interpolate observations with a one-to-one relationship. A constant negative bias from this relationship, however, serves as a measure of distance from the upper boundary.

4. Results and discussion

The $c_i(d)$ and $w(d)$ models and results are presented in sections 4.1 and 4.2 below and in Table 2. The final weight models for $c_i(d)$ and $w(d)$ generally corresponded to a lognormal residual distribution or one intermediate between the Poisson and lognormal (the GLS regressions required two to three iteration cycles on the weight model). The $c_i(d)$ models are given in Table 3 and analysis in section 4.3. Bounding statistics are listed in Tables 4 and 5 and analyses are listed in section 4.4.

4.1. Aerosol optical depth ($c_i$)

In all cases, the $c_i(d)$ curves (Table 2, section a) were best modeled with a simple sinusoid and the normal curves were nearly identical to NREL's corresponding monthly mean based model with the average parameter agreement being within a few percent (Appendix B, Table B-1, NREL, 1995). The worst regression and overall disagreement was for the Dodge City site.
The data scatter was comparatively much more than that for any other site, the constant was 6% smaller (the second largest difference for this parameter), the amplitude coefficient was 25% smaller, and the phase angle which was fit not fixed was 14% larger. The Ames normal model was compared to the one NREL listed for Des Moines (41.52° N, 93.65° W, and 294 m elevation (Appendix B, Table B1, NREL, 1995)]. The distance between Ames and Des Moines is about 60 km (= 0.54° latitude). The Ames amplitude parameter was 14% higher than the one for Des Moines, while the mean was 10% higher, and the phase angle date was 3% higher. Since parameter standard errors were not reported for any location, more formal statistical comparisons were not done. The agreement, however, seems reasonable considering that my normal model was developed differently from the ones listed in NREL's Appendix B, Table B1 (NREL, 1995). My daily median-based data selection procedure used in this study along with the weighted regressions should, in general, be a more robust methodology than NREL's. NREL used OLS on monthly means (Chapter 6, NREL, 1995). The climatic minimum curves all fit very well, and were developed similarly. Hence the GLS models developed in this work were utilized for $c_n(d)$ curves rather than NREL's models. Typical results, as shown in Fig. 1, have the climatic normal distinct from the climatic minimum throughout the annual cycle (P<0.05). The occasional overlapping of the 95% confidence limits for some periods occurred within the annual cycle for both the $c_n$ and $w$ comparisons. In all such cases, however, the curves can be made distinct throughout the annual cycle by relaxing the P level, generally to P = 0.1.

Error differences (sometime called "sensitivity") were not constant throughout the year (they were largest at or near the summer solstice and smallest in the winter). Model 2 differences were comparatively larger than those for the corresponding model 1 results. On average a decrease of 0.01 from the normal $c_n$ increased the average daily solar radiation estimate by 0.09 MJ m$^{-2}$ in model 1 and 0.35 MJ m$^{-2}$ in model 2; so model 2 is about 4 times as sensitive to changes in $c_n$ as model 1! This result is
difficult to show analytically because \( \frac{d(model \ 1)}{\partial c_n} \) is a highly nonlinear expression but \( \frac{d(model \ 1)}{\partial c_n} / \frac{d(model \ 2)}{\partial c_n} \) can be evaluated on any time scale and over any period numerically. An analytic expression for the absolute value of this ratio was developed for use with the Ames normal model. In a given day the value of the hourly ratio varied throughout the day generally peaking at noon with an average value of 0.285 but ranged from 0.069 excluding sunrise and sunset hours. The daily value was not as varied throughout the year and had a median value of 0.259 which is about 0.09/0.35. More complicated sensitivity analyses for all inputs were conducted but revealed little additional insight. Errors in the component inputs were obviously additive.

4.2. Precipitable water (w)

Gaussian regression models for the normal and minimum w trends were selected because in each case they interpolated the data better than sine curves (Table 2, section b), as the Wooster curves and data show (Fig. 2). Again, generally the climatic normals were distinct from the climatic lows (P<0.05).

As with the \( c_n \) errors, prediction differences were not constant throughout the year (they were largest during early spring and smallest during midsummer) and model 2 differences were somewhat larger than those for the corresponding model 1 results. The magnitude of these differences was smaller than any of those for \( c_n(d) \). A decrease of 1 mm from the normal w(d) increased the average daily solar radiation estimate by 0.04 MJ m\(^{-2}\) in model 1 and 0.05 MJ m\(^{-2}\) in model 2.

4.3. Surface albedo (\( \alpha \))

Although they had a slight lack of fit in the early part of the year for both sites, splined polynomial regression models were selected as the best means to interpolate the \( \alpha \) data. The Ames curve and data were typical (Fig. 3). Based on normal snowfall, air temperature, and cropping patterns
for central Iowa, the Ames model and data seemed reasonable (Waite and Hillaker, 1982; ISU, 1986).

Error analysis results were similar to those for the $\alpha_d$ input but smaller in magnitude making $\alpha$ the least consequential component in the models. An increase of 0.1 from the normal $\alpha(d)$ increased the average daily solar radiation estimate by 0.20 MJ m$^{-2}$ in model 1 and 0.13 MJ m$^{-2}$ in model 2.

4.4. Clear day solar radiation ($SI_c$)

For each site, the selection procedure of each maximum daily $SI_c$ datum from the 30 year record does not necessarily yield an observation that lies on the climatic clear day boundary. While it is not known what the relevant ambient meteorological conditions were for data that the selection process chose for the Ames and Wooster sites, the ambient conditions were known for the chosen $SI_c$, $w$, and $\sigma_s$ data from the SAMSON sites. The SAMSON site $SI_c$ data were generally but not exclusively from very low total sky cover (tsc) days. For the Bismarck, Columbia, and Dodge City SAMSON sites collectively, 51% of the selected data had a tsc < 0.05, 65% < 0.10, 82% < 0.20, and 90% < 0.30. The correlation between tsc and $SI_c$ was insignificant at any reasonable probability level. A common date for the selection of maximum $SI_c$, minimum $w$, and minimum $\sigma_s$ was rare. Bismarck had 10 such observations, the other two had 22 each. Collectively these total 4.9% of the selected data. This subset is designated the "reselected" data. It is admittedly not an independent confirmation set and has the same kinds of problems but hopefully to a lesser extent. In all of the sites the selected $SI_c$ data were skewed in time toward the first two decades of the period of record. Ames and Wooster, however, were comparatively much more so. For the Bismarck, Columbia, and Dodge City SAMSON sites collectively, 63% of the data were drawn from the first half of the period of record and 79% from the first two decades. The corresponding values for Ames were 89% and 97%; those for Wooster were 75%
and 87%. In the Ames selected data, over a week of sequential observations from the mid 1960's exceeded extraterrestrial values and so were replaced with the next highest observation. Wooster only had three such points, hence they were just deleted. The SI_e data selection was repeated for the Akron/Canton secondary SAMSON site and compared to the set for Wooster. The Wooster data were on average 1.25 MJ m^-2 higher. Comparing the scatter plots for the chosen SAMSON sites to those from Ames and Wooster reveals the latter are much more scattered and probably not of the same quality. Data outside of the SAMSON database are likely not to be of the same quality.

Both broadband models using the climatic normal inputs underestimated the boundary for the selected observations. Model 1 estimates generally interpolated the data while the model 2 estimates greatly underestimated the data (Table 4). The reason is probably because model 2 is much more sensitive to S_e. Hence the use of either model, especially model 2, with climatic normals is unsatisfactory. Perhaps data that fall below the model 1 curve should be excluded since the selection procedure does not guarantee that a datum is truly in the clear day upper boundary population. For the Bismarck, Columbia, and Dodge City SAMSON sites collectively, the selected observations were bound by both the climatic minimum and LL1 inputs although a few of the selected data were above the model 2 estimates from minimums for Columbia and Dodge City (Table 4). Results for Columbia were typical (Fig. 4). For Ames and Wooster the assessment was more difficult because of the greater scatter and probable lower quality of the data; however, the prevailing patterns are the same (Table 4). The data show that all of the observations that lie above the model 1 boundary for both the Ames and Wooster sites were highly skewed to the earlier part of the observation period. For Ames 99.5% of the outliers were in the first two decades and 97.3% in the first 15 years; the comparative values for Wooster were 98.0% and 95.0%. Perhaps before the automation of the data acquisition and processing process, the methodology favored over-estimation
especially on the clearer days. If so, the Ames and Wooster data could be adjusted to become more consistent with the rest. For all sites, the magnitude of the model/data difference was positively correlated with $C_{L}$ (not constant throughout the annual period). The relative value of the differences behaved in the opposite way but comparatively the effect was greatly reduced. Based on climatic minimum in $w$ and $C_{L}$, model 1 relative differences, the data for the SAMSON primary sites were bound on average by 6.2% (1.3 MJ m$^{-2}$) and for model 2 by 5.8% (1.1 MJ m$^{-2}$); the corresponding values for the LL1 conditions were 7.8% (1.7 MJ m$^{-2}$) for model 1 and 8.5% (1.8 MJ m$^{-2}$) for model 2.

There is no general agreement or precedent on selecting the best model for bounding $C_{L}$ data. Furthermore, while the long-term data come from high precision radiometers, there are many possible problems with using such data for modeling and model confirmation (e.g., see Atwater and Ball (1978a) which included radiometer calibration and drift problems). Both models and data are always questionable. Over the long run the very best data, including even the reselected data, are probably not better ± 1 MJ m$^{-2}$ or 5%. So ±1 MJ m$^{-2}$ is proposed as minimum condition for bounding the selected data provided the boundary curve is otherwise satisfactory. Still a choice is not clear-cut and cannot be based on only one or two performance measures. Instead a collective assessment should be made based on scatter plots, residual plots, and several different types of performance measures using both the selected data but particularly the reselected data because many of the selected data were probably not as good candidates for this task. For this analysis, both models employed the climatic low inputs. Based on bias alone for the selected data, model 1 always exceeded the proposed value but model 2 did not for one case, Columbia. Based on bias alone for the reselected data both models failed and each bias value was statistically different from 1 MJ m$^{-2}$ (P<0.05). Using all error measures, however, model 1 was reasonably close to +1 MJ m$^{-2}$ while model 2 was not (Table 5 and Fig. 5). Other means of assessment
considered were similar and comparatively very close (Table 5); in general, they favored model 1 but not to the exclusion of model 2. Given its greater simplicity, model 2 is a sound practical choice. If one desires to be more conservative, however, model 2 should probably be used with the LL1 inputs.

Since both models are run on an hourly basis, conceptually they could be used for screening hourly SI data. A test plan for this job needs to be developed and carried out first.

Finally, in practice, a metamodel for the SI estimates for a given site could be developed and used in place of the SI model. In this case a 'metamodel' is a regression of the broadband estimates on day in the year. Metamodels could then be used when computational simplicity, time, or costs are important.

5. Conclusions

When normal trend estimates of $c_a$ and $w$ for each of two long-term radiometer stations were used to predict SI with either one of two broadband solar radiation models, a systematic under-estimation of the upper boundary resulted in each case with model 1 generally interpolating the selected data and model 2 generally completely underestimating them. Employing input component models developed from the climatic minima data for each site completely bounded all the selected data with model 1 and almost did so with model 2 but only for the chosen SAMSON sites. The Ames and Wooster data, which are probably systematically high and of lower quality, had similar patterns but had about half their observations above the boundary. Employing the lower confidence limits from the latter climatic input models (LL1 inputs) gave a more conservative boundary for both models at all sites.

The most influential component in both models was $c_a$ while $a$ was the least. For practical purposes model 2, which is simpler, worked as well as model 1 for bounding the data. The methodology can be used for better
estimation of $SI_{s}$ at locations throughout the United States that lack long-term radiometer data (e.g., many of the locations in the Cooperative Station Network, hydrological studies, etc.).

Acknowledgments

This work was supported by the USDA-ARS-MWA National Soil Tilth Laboratory, Dr. J.L. Hatfield, Director. Prof. R.E. Carlson, ISU, Ames, IA, provided the Ames solar radiation data. Profs. J. Holeman and D. Elwell, OSU, Columbus, OH provided the Wooster solar radiation data and meteorological information. Prof. D.N. Yarger provided advice, encouragement, and comments on the manuscript. Prof. W.D. Sellers, UA, Tucson, AZ, provided comments on the manuscript. Dr. E.L. Maxwell, USDE-NREL, Golden, CO, provided comments on the manuscript and information on SAMSON. Anonymous reviewers also provided comments, advice and assistance. Mrs. J. Meek graciously edited the manuscript.

Appendix

Definition of Symbols

$a$ surface albedo (dimensionless fraction)

$a_w$ water vapor absorptance for model 2 (dimensionless fraction)

d day in the year (integer day number, 1 to 365)

$F_s$ fraction of forward to total scattering (dimensionless)

$I_0$ solar constant in an hourly model (4.921 MJ m$^{-2}$ [1367 W m$^{-2}$])

$l_o$ Ozone layer thickness (cm)

$m$ relative optical air mass (dimensionless)

$m_{o}$ absolute (pressure corrected) optical air mass

$m_o$ ozone relative optical air mass

$n$ number of observations in a data set

$P$ atmospheric pressure (kPa)

$P_o$ standard atmospheric pressure (101.325 kPa)

$r_s$ sky albedo (dimensionless fraction)

$SI$ shortwave solar irradiance (MJ m$^{-2}$)
S_{c}  clear day shortwave solar irradiance (MJ m^{-2})
\sigma_a  broadband aerosol optical depth (dimensionless or m^{-1})
\sigma_{\text{anom}}  normal trend estimate of \sigma_a
\sigma_{\text{min}}  minimum trend estimate of \sigma_a
\sigma_{\text{all}}  95% lower confidence limit for \sigma_{\text{min}}
T_a  transmittance of aerosols (dimensionless fraction)
T_{aa}  transmittance of aerosol absorptance (dimensionless fraction)
T_{as}  transmittance of aerosol scattering (dimensionless fraction)
T_f  transmittance of uniformly mixed gases (dimensionless fraction)
T_m  transmittance (global) of all molecular effects except water vapor for model 2 (dimensionless fraction)
T_{\text{Linke}}  Linke turbidity factor (Rayleigh atmospheres)
T_R  transmittance of Rayleigh scattering (dimensionless fraction)
T_o  transmittance of ozone absorptance (dimensionless fraction)
T_w  transmittance of water vapor absorptance (dimensionless fraction)
U_o  total optical path length for ozone (cm)
U_w  total optical path length for water vapor (cm)
w  precipitable water thickness (cm)
w_{\text{anom}}  normal trend estimate of w
w_{\text{min}}  minimum trend estimate of w
w_{\text{all}}  95% lower confidence limit for w_{\text{min}}
w_o  single-scattering albedo (dimensionless fraction)
Z  solar zenith angle (degrees or radians depending on the equation)

References


Fig. 1. Adjusted normal (upper set, solid square) and minimum (lower set, solid circle) aerosol depth data and annual trend models with 95% confidence limits for Ames, IA. Data were for the period 1961-1990. Regression results are listed in Table 2.
Fig. 2. Normal (upper set, solid square) and minimum (lower set, solid circle) precipitable water data and annual trend curves with 95% confidence limits for Akron/Canton, OH. Data were for the period 1961-1990. Regression results are listed in Table 2.
Fig. 3. Monthly mean albedo data and annual trend curve for Ames, IA. Data were for an unspecified period. Regression results are listed in Table 3.
Fig. 4. Clear day solar radiation model results for Columbia, MO, model 1 in (a) and model 2 in (b). In each figure the top solid line is for the extraterrestrial model, the upper broken line is for the LL1 inputs, the lower solid line is for the climatic low inputs, and the bottom broken line is for the climatic normal inputs. The solid circles are the selected clear day data.
Fig. 5. (a) Berg plot (1992) for the reselected observations and the model 1 predictions for the climatic low input components. Summary statistics are listed in Table 5. (b) Differences of the observed and predicted values shown in Fig. 5a but plotted against day in the year.
### Table 1

**Site geographic coordinates and data information.**

<table>
<thead>
<tr>
<th>Site, State</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation</th>
<th>Data Type</th>
<th>Source</th>
<th>Record Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames, IA</td>
<td>42.03° N</td>
<td>93.80° W</td>
<td>335.0 m</td>
<td>SI(d)</td>
<td>ISU</td>
<td>31 yr</td>
</tr>
<tr>
<td>Akron, OH</td>
<td>40.92° N</td>
<td>81.42° W</td>
<td>377.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Bismarck, ND</td>
<td>46.77° N</td>
<td>100.75° W</td>
<td>502.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Columbia, MO</td>
<td>38.82° N</td>
<td>92.22° W</td>
<td>270.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Dodge City, KS</td>
<td>37.77° N</td>
<td>99.97° W</td>
<td>787.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Madison, WI</td>
<td>43.12° N</td>
<td>89.32° W</td>
<td>262.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Omaha, NE</td>
<td>41.37° N</td>
<td>96.52° W</td>
<td>404.0 m</td>
<td>(\sigma_\alpha)</td>
<td>SAMSON</td>
<td>30 yr</td>
</tr>
<tr>
<td>Toronto(^a)</td>
<td>43.75° N</td>
<td>79.50° W</td>
<td>178.3 m</td>
<td>(\sigma_\alpha)</td>
<td>Iqbal, 1983</td>
<td>NA</td>
</tr>
<tr>
<td>Wooster, OH</td>
<td>40.78° N</td>
<td>81.92° W</td>
<td>310.9 m</td>
<td>SI(d)</td>
<td>OSU</td>
<td>31 yr</td>
</tr>
</tbody>
</table>

\(^a\) SI is shortwave irradiance (MJ m\(^{-2}\)), \(\sigma_\alpha\) is aerosol optical depth (dimensionless), \(w\) is precipitable water (cm), and \(\alpha\) is albedo.

\(^b\) ISU (Iowa State University); SAMSON (Solar And Meteorological Observation Network); and OSU (Ohio State University, a.k.a. OARDC).

\(^c\) NA is not available.

\(^d\) Canada
### Table 2
Regression component models tested in the broadband clear day routine.

<table>
<thead>
<tr>
<th>Site</th>
<th>n</th>
<th>$R^2$</th>
<th>Model</th>
<th>Weight Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aerosol optical depth models, $a_*$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akron/Canton</td>
<td>365</td>
<td>0.80</td>
<td>median: $a_{\text{median}}(d) = 0.123 + 0.042 \cos((2\pi365)^{-1}(d - 183))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>0.93</td>
<td>minimum: $a_{\text{min}}(d) = 0.060 + 0.021 \cos((2\pi365)^{-1}(d - 184))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td>Ames</td>
<td>365</td>
<td>0.98</td>
<td>median: $a_{\text{median}}(d) = 0.110 + 0.051 \cos((2\pi365)^{-1}(d - 189))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>0.86</td>
<td>minimum: $a_{\text{min}}(d) = 0.062 + 0.030 \cos((2\pi365)^{-1}(d - 191))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td>Bismarck</td>
<td>365</td>
<td>0.96</td>
<td>median: $a_{\text{median}}(d) = 0.071 + 0.030 \cos((2\pi365)^{-1}(d - 185))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>0.83</td>
<td>minimum: $a_{\text{min}}(d) = 0.029 + 0.012 \cos((2\pi365)^{-1}(d - 187))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td>Columbia</td>
<td>360</td>
<td>0.97</td>
<td>median: $a_{\text{median}}(d) = 0.100 + 0.042 \cos((2\pi365)^{-1}(d - 187))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td></td>
<td>352</td>
<td>0.88</td>
<td>minimum: $a_{\text{min}}(d) = 0.043 + 0.020 \cos((2\pi365)^{-1}(d - 189))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td>Dodge City</td>
<td>357</td>
<td>0.30</td>
<td>median: $a_{\text{median}}(d) = 0.072 + 0.031 \cos((2\pi365)^{-1}(d - 208))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>0.91</td>
<td>minimum: $a_{\text{min}}(d) = 0.031 + 0.017 \cos((2\pi365)^{-1}(d - 191))$</td>
<td>$s^2 = \hat{y}^2$</td>
</tr>
</tbody>
</table>

| **Precipitable water models, $w$**     |     |       |       |              |
| Akron/Canton    | 365 | 0.98  | median: $w_{\text{median}}(d) = 0.67 + 2.20 \exp(-0.000144(d - 207)^2)$ cm | $s^2 = \hat{y}^2$ |
|                 | 365 | 0.96  | minimum: $w_{\text{min}}(d) = 0.35 + 1.42 \exp(-0.000138(d - 208)^2)$ cm | $s^2 = \hat{y}^2$ |
| Ames            | 365 | 0.98  | median: $w_{\text{median}}(d) = 0.64 + 2.48 \exp(-0.000145(d - 204)^2)$ cm | $s^2 = \hat{y}^2$ |
|                 | 365 | 0.92  | minimum: $w_{\text{min}}(d) = 0.27 + 1.53 \exp(-0.000157(d - 204)^2)$ cm | $s^2 = \hat{y}^2$ |
| Bismarck        | 365 | 0.98  | median: $w_{\text{median}}(d) = 0.47 + 1.88 \exp(-0.000154(d - 204)^2)$ cm | $s^2 = \hat{y}^2$ |
|                 | 365 | 0.94  | minimum: $w_{\text{min}}(d) = 0.17 + 1.19 \exp(-0.000198(d - 205)^2)$ cm | $s^2 = \hat{y}^2$ |
| Columbia        | 365 | 0.96  | median: $w_{\text{median}}(d) = 0.72 + 2.73 \exp(-0.000141(d - 204)^2)$ cm | $s^2 = \hat{y}^2$ |
|                 | 362 | 0.87  | minimum: $w_{\text{min}}(d) = 0.25 + 1.43 \exp(-0.000155(d - 203)^2)$ cm | $s^2 = \hat{y}^2$ |
| Dodge City      | 365 | 0.98  | median: $w_{\text{median}}(d) = 0.61 + 2.30 \exp(-0.000163(d - 205)^2)$ cm | $s^2 = \hat{y}^2$ |
|                 | 365 | 0.91  | minimum: $w_{\text{min}}(d) = 0.27 + 1.41 \exp(-0.000195(d - 207)^2)$ cm | $s^2 = \hat{y}^2$ |

* Most data sets had at least one outlier. In the regressions, observations with weighted residuals that were outliers at the $P<0.001$ level were removed when there were three or more and then the regression was redone.

b All parameters have $P<0.01$ or better based on their $T$ value.

c The weight models are for inverse variance weighting, i.e., weight=$1/s^2$. Here $s^2$ is the variance and $\hat{y}$ is the predicted value.
<table>
<thead>
<tr>
<th>Site</th>
<th>n</th>
<th>R²</th>
<th>Model²</th>
</tr>
</thead>
</table>
| Ames       | 12| 0.997 | \[ a(d) = \begin{cases} 
0.612 + 2.45 \times 10^{-3}d - 5.60 \times 10^{-5}d^2, & d \leq 116 \\
0.140, & 116 < d \leq 284 \\
0.140 + 0.0112(d-284) - 5.92 \times 10^{-5}(d - 284)^2, & d > 284 
\end{cases} \] |
| Bismarck   | 12| 0.995 | \[ a(d) = \begin{cases} 
0.660, & d \leq 74 \\
0.660 - 0.0608(d-74), & 74 < d \leq 153 \\
0.180, & 153 < d \leq 256 \\
0.180 + 0.0927(d-256) - 4.30 \times 10^{-5}(d - 256)^2, & d > 256 
\end{cases} \] |
| Columbia   | 12| 0.999 | \[ a(d) = \begin{cases} 
0.596 + 1.19 \times 10^{-3}d - 4.76 \times 10^{-5}d^2, & d \leq 111 \\
0.140, & 111 < d \leq 287 \\
0.140 + 9.56 \times 10^{-3}(d-287) - 4.68 \times 10^{-4}(d - 287)^2, & d > 287 
\end{cases} \] |
| Dodge City | 12| 0.996 | \[ a(d) = \begin{cases} 
0.651 - 3.29 \times 10^{-3}d, & d \leq 120 \\
0.180, & 120 < d \leq 285 \\
0.180 + 0.0103(d-285) - 5.16 \times 10^{-3}(d - 285)^2, & d > 285 
\end{cases} \] |
| Wooster    | 12| 0.999 | \[ a(d) = \begin{cases} 
0.537 + 1.38 \times 10^{-3}d - 3.73 \times 10^{-4}d^2, & d \leq 114 \\
0.213, & 114 < d \leq 284 \\
0.213 + 3.47 \times 10^{-3}(d - 284) - 5.52 \times 10^{-4}(d - 284)^2, & d > 284 
\end{cases} \] |

² All parameters have P≤0.1 or better based on their T value.
<table>
<thead>
<tr>
<th>Comparison Site</th>
<th>MAE</th>
<th>MBE</th>
<th>MBE Relative (%)</th>
<th>MAX</th>
<th>MIN</th>
<th>Predicted&lt;Observed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normals Ames</td>
<td>1.12</td>
<td>-0.88±0.06</td>
<td>5.71±0.40</td>
<td>0.75 (58)</td>
<td>-31.73 (347)</td>
<td>78.5</td>
</tr>
<tr>
<td>Bismarck</td>
<td>0.56</td>
<td>0.51±0.02</td>
<td>2.76±0.13</td>
<td>11.19 (323)</td>
<td>-3.65 (5)</td>
<td>10.7</td>
</tr>
<tr>
<td>Columbia</td>
<td>0.43</td>
<td>0.17±0.03</td>
<td>1.32±0.14</td>
<td>10.53 (362)</td>
<td>-5.41 (142)</td>
<td>33.4</td>
</tr>
<tr>
<td>Dodge City</td>
<td>0.44</td>
<td>0.32±0.02</td>
<td>1.73±0.11</td>
<td>8.08 (338)</td>
<td>-3.44 (132)</td>
<td>21.4</td>
</tr>
<tr>
<td>Wooster</td>
<td>1.10</td>
<td>0.00±0.07</td>
<td>1.35±0.27</td>
<td>15.01 (357)</td>
<td>-40.45 (356)</td>
<td>47.5</td>
</tr>
<tr>
<td><strong>Minima</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ames</td>
<td>0.99</td>
<td>-0.11±0.06</td>
<td>1.96±0.39</td>
<td>10.84 (182)</td>
<td>-27.07 (347)</td>
<td>50.8</td>
</tr>
<tr>
<td>Bismarck</td>
<td>1.39</td>
<td>1.39±0.04</td>
<td>6.95±0.13</td>
<td>15.33 (323)</td>
<td>0.90 (5)</td>
<td>0</td>
</tr>
<tr>
<td>Columbia</td>
<td>1.34</td>
<td>1.34±0.03</td>
<td>6.31±0.13</td>
<td>14.83 (362)</td>
<td>0.11 (142)</td>
<td>0</td>
</tr>
<tr>
<td>Dodge City</td>
<td>1.20</td>
<td>1.20±0.02</td>
<td>5.43±0.11</td>
<td>11.82 (338)</td>
<td>0.10 (84)</td>
<td>0</td>
</tr>
<tr>
<td>Wooster</td>
<td>1.43</td>
<td>0.80±0.08</td>
<td>2.30±0.43</td>
<td>18.45 (357)</td>
<td>-34.76 (356)</td>
<td>27.6</td>
</tr>
<tr>
<td><strong>LL1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ames</td>
<td>1.08</td>
<td>0.25±0.07</td>
<td>-0.36±0.39</td>
<td>12.45 (182)</td>
<td>-25.24 (347)</td>
<td>40.1</td>
</tr>
<tr>
<td>Bismarck</td>
<td>1.71</td>
<td>1.71±0.04</td>
<td>8.49±0.13</td>
<td>16.98 (323)</td>
<td>2.85 (130)</td>
<td>0</td>
</tr>
<tr>
<td>Columbia</td>
<td>1.78</td>
<td>1.78±0.03</td>
<td>7.99±0.12</td>
<td>16.00 (362)</td>
<td>2.18 (142)</td>
<td>0</td>
</tr>
<tr>
<td>Dodge City</td>
<td>1.54</td>
<td>1.54±0.03</td>
<td>6.79±0.11</td>
<td>13.13 (338)</td>
<td>1.52 (89)</td>
<td>0</td>
</tr>
<tr>
<td>Wooster</td>
<td>1.56</td>
<td>1.00±0.08</td>
<td>3.21±0.42</td>
<td>19.31 (357)</td>
<td>-33.35 (356)</td>
<td>24.9</td>
</tr>
</tbody>
</table>

| **Model 2**     |     |     |                  |     |     |                       |
| Normals Ames    | 2.58| -2.57±0.06 | -14.69±0.42 | 1.65 (58) | -42.39 (347) | 99.7 |
| Bismarck        | 0.84| -0.81±0.04 | -3.97±0.14  | 4.61 (323)  | -10.32 (134) | 92.3 |
| Columbia        | 1.86| -1.85±0.06 | 8.54±0.21   | 3.09 (362)  | -18.33 (174) | 98.6 |
| Dodge City      | 0.99| -0.91±0.04 | -3.81±0.17  | 3.34 (86)   | -10.67 (237) | 84.7 |
| Wooster         | 1.95| -1.90±0.07 | -11.29±0.49 | 7.22 (64)   | -54.35 (356) | 93.9 |
| **Minima**      |     |     |                  |     |     |                       |
| Ames            | 0.98| -0.82±0.06 | -5.17±0.38  | 9.09 (58)   | -30.18 (347) | 76.2 |
| Bismarck        | 1.34| 1.34±0.03  | 7.08±0.14   | 16.00 (323) | 1.03 (134) | 0 |
| Columbia        | 0.96| 0.94±0.02  | 4.96±0.16   | 15.05 (362) | -2.68 (174) | 3.6 |
| Dodge City      | 1.11| 1.11±0.02  | 5.39±0.14   | 13.02 (338) | -0.32 (132) | 0.8 |
| Wooster         | 1.10| 0.12±0.07  | -0.69±0.43  | 16.43 (357) | -38.13 (356) | 44.2 |
| **LL1**         |     |     |                  |     |     |                       |
| Ames            | 0.94| -0.08±0.06 | -1.62±0.37  | 11.63 (58)  | -25.74 (347) | 50.8 |
| Bismarck        | 1.92| 1.92±0.04  | 9.72±0.14   | 17.84 (323) | 3.61 (134) | 0 |
| Columbia        | 1.76| 1.76±0.03  | 8.05±0.13   | 16.81 (362) | 1.88 (142) | 0 |
| Dodge City      | 1.69| 1.69±0.02  | 7.66±0.14   | 15.06 (338) | 2.10 (138) | 0 |
| Wooster         | 1.25| -0.51±0.08 | 1.13±0.43   | 18.03 (357) | -35.49 (356) | 32.9 |

*aTerminology: MAE: Median bias error; MBE: Mean bias error; MAX: Maximum (value in parentheses is the day in the year the maximum was realized); MIN: Minimum (value in parentheses is the day in the year the minimum was realized)

*b Comparison Parameterizations (for 365 observations at all but Wooster which had 362)

Normals: $S_{1m}(d) - S_{L} (\sigma_{m}(d), w_{m}(d), \omega(d))$ in MJ m$^{-2}$; Minima: $S_{1min}(d) - S_{L} (\sigma_{m}(d), w_{m}(d), \omega(d))$ in MJ m$^{-2}$; LL1: $S_{L}(d) - S_{L} (\sigma_{m}(d), w_{m}(d), \omega(d))$ in MJ m$^{-2}$
Table 5
Quantitative measures of clear day shortwave irradiance model performance. Model inputs are for climatic minima in aerosol optical depth and precipitable water*. Observations are selected from all sites.

<table>
<thead>
<tr>
<th>Model</th>
<th>N</th>
<th>D</th>
<th>P</th>
<th>Ratio</th>
<th>r</th>
<th>MAE</th>
<th>RMSE</th>
<th>Bias</th>
<th>$s^2_b$</th>
<th>$b_0$</th>
<th>$b_1$</th>
<th>SEE</th>
<th>PRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>20.92</td>
<td>21.75</td>
<td>0.962</td>
<td>0.999 (P&lt;0.001)</td>
<td>0.826</td>
<td>0.939</td>
<td>-0.826±0.061</td>
<td>-0.501±1.174</td>
<td>0.984±0.008</td>
<td>0.44</td>
<td>10.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>54</td>
<td>20.92</td>
<td>21.60</td>
<td>0.969</td>
<td>0.998 (P&lt;0.001)</td>
<td>0.727</td>
<td>0.849</td>
<td>-0.681±0.070</td>
<td>-1.021±2.04</td>
<td>1.016±0.009</td>
<td>0.50</td>
<td>14.23</td>
<td></td>
</tr>
</tbody>
</table>

**Terminology**

N: Number of observations.

O: Mean observed value of maximum clear day measurements $S_{I_{max}}(d)$ in MJ m$^{-2}$.

P: Mean predicted value of maximum clear day measurements $S_{I_{max}}(d)$ in MJ m$^{-2}$.

Ratio: (Mean observed)/(Mean predicted)

r: Pearson product-moment correlation coefficient.

MAE: Mean absolute error (MJ m$^{-2}$).

RMSE: Root Mean square error (MJ m$^{-2}$).

Bias: Mean bias error (MJ m$^{-2}$).

$s^2_b$: Bias variance.

$b_0$: Intercept estimate in the ordinary least squares regression model $S_{I_{max}}(d) = b_0 + b_1 S_l(d)$, with the standard error of the parameter estimate.

$b_1$: Line 1: Slope estimate in the ordinary least squares regression model $S_{I_{max}}(d) = b_0 + b_1 S_l(d)$, with the standard error of the parameter estimate.

Line 2: Slope only estimate in the ordinary least squares regression model $S_{I_{max}}(d) = b_1 S_l(d)$, with the standard error of the parameter estimate.

SEE: Standard error of the regression estimate (MJ m$^{-2}$).

PRESS: Prediction error sum of squares.
AREAL SCALES OF FLUCTUATION FOR IOWA'S ALBEDO IN 1990

A paper prepared for Water Resources Research

David Meek

Abstract

Understanding spatial variability in surface albedo is pertinent to understanding spatial variation in surface net radiation, surface energy balance, and hence evapotranspiration. Related research has focused on aggregation methods, assuming the regions or subregions of aggregation are known. This study was conducted to estimate and examine areal scales of fluctuation to determine a level of aggregation of surface albedo for the state of Iowa throughout the 1990 growing season for 16 two-week periods in 1990 from March through early October. NOAA-11 AVHRR data were used to construct georegistered albedo for 1 km² areas. For each time period, two directional semivariograms were developed for both eastings and northings. The first used the albedo data without regard to possible trend contamination. The second used detrended data, specifically residuals from a median polish that removed large scale spatial trends. Directional scales of fluctuation were estimated with an iterative numerical integration of the correlogram associated with each semivariogram. Areal scales of fluctuation were estimated from the directional values. Results from the raw data were often very different from the ones developed from detrended data. In some cases the results from the raw data were questionable and suggest a scale of variation beyond the half-length limit of the development sets. For the detrended data, results seemed reasonable for all but the last period when harvest was occurring. Areal scales of fluctuation were not constant throughout the growing season and have a lag-1 autocorrelation structure. Values ranged from about 300 to 2000 km². They were smallest during the maturing stages of the major crops. Thus ambient conditions affect surface albedo and could be used to selectively sample other albedo data sets. Averaging areas probably need to be at least double that of the areal scale of fluctuation. Although the findings...
are preliminary, indications are that climate modelers can consider all of Iowa to be homogeneous in albedo at each crop developmental period in the growing season and can readily average over the whole state. Researchers working closer to the point scale in real time may have to more carefully assess local variability in light of their goals and available data on the site. Comparison of the results in this study with those from other aggregation concepts as well as those from additional data sets representing other scales, years, and environments should be considered.

1. Introduction

Surface albedo (total hemispherical shortwave reflectance at the earth's surface) is a fundamental component of the surface net radiation and therefore the surface energy balance. Considerable research has focused on methodology for areal and regional estimation of the surface energy balance, especially for modeling evapotranspiration from a scaling-up approach [see e.g., Kirby, 1996; Sellers et al., 1996; Henderson-Sellers et al., 1995; Henderson-Sellers et al., 1993; Bolle, 1993; Black et al., 1989]. In view of this interest in relation to energy balance research, it is not surprising that considerable effort has focused on surface-albedo estimation from areal to planetary spatial scales [see e.g., Norman et al., 1996; Sellers et al., 1996; Henderson-Sellers et al., 1993; Ranson et al., 1991; Houghton et al., 1990; Dedieu et al., 1987]. In addition, Sagan et al. [1979] discussed anthropogenic changes in surface albedo over time, including the span of mankind's existence.

Many data intensive field experiments, e.g., MONSOON '90 [Kustas and Goodrich, 1994], have planned to use remote-sensing data along with point-source data in a scaling-up approach to assessing areal energy balance components. Areal scales of interest vary from watershed or subwatershed areas (<100 km²) to general circulation model [GCM] grid cell size [generally either 10⁴ km² or 1.6×10⁵ km², see e.g., Schmugge and André, 1991 or Sellers et al., 1996]. Programs like HAPEX-MOBILHY [Hydrologic
Atmospheric Experiment - Modélisation du Bilan Hydrique] were for CGM grids while projects like FIFE [the First ISLSCP (the International Satellite Land Surface Climatology Project) Field Experiment, =225 km²] fell in between the stated scale limits. Some researchers, including Short et al. [1993], Shuttleworth [1988], and Dooge [1986], have argued that, although difficult, there is a need to reformulate the basic processes to suit the scale of interest. Though developing methodology for scaling-up is abundant, there is no standard and numerous approaches exist [e.g., see Braden, 1995; Famiglietti and Wood, 1995; Hashmi et al., 1995; Hofstee et al., 1993; or Lhomme, 1992]. Hence it is no surprise that there is much discussion of inherent problems, including many that involve scaling and/or aggregation issues [e.g., see Kirby, 1996; Blöschl and Sivapalan, 1995; Henderson-Sellers et al., 1995; Bolle, 1993; Becker, 1989; or Hatfield, 1985]. Raupach [1993] discussed three tasks needed to scale land surface processes. The first of the three tasks listed was given as the nontrivial need to specify the surface characteristics, their variability, and their averaging areas within an entire heterogeneous domain of consideration. The importance of this task was stressed but no solution was offered. The remaining tasks concerned aggregation or disaggregation schemes.

The concept of aggregating data is inherently involved in using any spatial and temporal data. Over a decade ago, Rodríguez-Iturbe [1986] advocated a suitable methodological precedent. Using Vanmarcke's theory [1983] on random fields, one can compute averaging areas and times with an index called the scale of fluctuation. Rodríguez-Iturbe's [1986] work was based on theoretical rainfall models but Rodríguez-Iturbe [1986] and Cressie [1991] suggested the need to apply this spatial analysis methodology to other kinds of hydrological studies that have measurements rather than simulated data.

Surface albedo data covering 1 km² sections within Iowa were constructed from remotely sensed data [NOAA-11 AVHRR data] taken over 16 two-week periods in 1990. The objectives of this work were to estimate
spatial and temporal scales of fluctuation from directional semivariograms developed from the albedo data under stationary and nonstationary assumptions, and then to assess the results in the context of modeling regional energy balances using Rodríguez-Iturbe’s [1986] general recommendation for case studies.

2. Methods

In assessing the albedo data, the nature of the land and its uses are important, as well as the climate and specific practices during the period of observation. Summaries on Iowa’s general geography and climate, conditions in 1990, data acquisition and processing, along with the methods of analyses follow.

2.1. Iowa’s Geography and Climate

Iowa is in the West North Central part of the United States of America bounded on the east by the Mississippi River and mostly bounded on the west by the Missouri River. The State of Missouri forms the entire southern border between the two rivers. The State of Minnesota forms almost all the northern border with the State of South Dakota occupying the remaining section along the northwestern sector. Iowa roughly forms a rectangle with its larger dimension, about 515 km, being in the east to west direction approximately between 90° 30' W and 96° 30' W and the smaller dimension, about 338 km, being in the north to south direction approximately between 40° 35' N and 43° 30' N [see e.g., Rand McNally, 1993]. The state’s total area is only about 0.6% in water body surfaces [Pharos Books, 1990]. The mean elevation is about 335 m above sea level. The highest point is 509 m on the top of the Ocheyedan Mound in the northwestern part of the state and near the Minnesota border and the lowest point is 146 m in the southeast at the Mississippi River [Salisbury and Rafferty, 1995]. Generally elevation increases to the west and north.
The state's terrain is generally flat or gradually rolling, one of the many consequences of glaciation during the Ice Age. Landscapes, however, are diverse, being classified into seven topographic regions [Fig. 1, based on Prior, 1991]. The regions are as follows [the percentages are percent of total state land area, which is 145,698.5 km², based on data and classifications in the Iowa Geological Survey's topographic database, S. Hoyer, personal communication, 1996]: the Northwest Iowa Plains [8.2%], the Des Moines Lobe [21.1%], the Iowan Surface [16.6%], the Paleozoic Plateau [4.5%], the Southern Iowa Drift Plain [43.7%], the Loess Hills [2.4%], and the Alluvial Plains [3.5%]. There are two alluvial plains. One is along the southern section of the Mississippi River [1.5%] while the other is along the southern section of the Missouri River [2.0%]. In the northeast, the Paleozoic Plateau (a.k.a. the "Driftless Region"), hills often rise over 100 m in elevation from the Mississippi River and its tributaries which have made deep cuts into the land surface. The Des Moines Lobe is the flattest region of Iowa and lies in the north central to western part of the state. This region was formed by the planing and subsequent melting of the last major ice sheet in the area which, in part, formed Iowa's Great Lakes: Lake Okoboji, Clear Lake, Storm Lake, and Spirit Lake. It has some of the most fertile soil in the country. The Southern Iowa Drift Plain (a.k.a. the Dissected Till Plains), a rolling land form having older, more weathered and aged glacial till deposits, makes up most of the remaining land area in the state. There is a small area of flat till plains in the southeast. Till plains usually form rich soils.

Given its suitable physical geography and low population, agriculture plays an important role in Iowa's economy and dominates its land usage practices. Farms usually occupy about 86% of the land area with crop production the major operation although there is considerable livestock production. Corn and soybean are the major crops but hay, oat, alfalfa, and a few other crops are also important. Forested area is about 5%.
Iowa has a continental, temperate climate. Generally summers are hot and moist while winters are cold and dry. Annual average normal temperature ranges from 11 °C in the south to 8 °C in the north. Annual total precipitation ranges from 625 mm in the northwest to 850 mm in the southeast. Most precipitation usually falls in the spring and summer. Summertime thunderstorms often occur. Winter snowstorms and occasional blizzards are less frequent. Long periods of low available soil moisture and low rainfall are rare.

2.2. Climate and crop conditions in 1990

Overall the year 1990 was warmer than normal, especially in the winter months, although it was slightly cooler in the spring months and relatively mild in the summer [Fig. 2; data for the figure and for the information that follows are from USDC-NOAA et al., 1990]. Precipitation was much higher than normal; the 1990 total rainfall amount qualified the year as the ninth wettest on record up to the time. The March total of 110 mm was a record value for the month and the June total of 204 mm was a near record value. Snowfall was less than normal due to the warm 1989-1990 winter. By March there were only residual patches of snow even after a midmonth snowstorm, which occurred mainly in the western half of the state. Thereafter and throughout the period of satellite observations there was no further significant snowfall or snow accumulation. The next snowfall was past the middle of October. These factors taken together made the year unusual climatically but in such a way that crop growth and development was nearly optimal although somewhat delayed. Overall the topsoil moisture was rarely classified as short at times that could harm the crops and then only for brief periods at a few localities. In April through the first week in May, top-soil moisture was classified as short in about 23% of the state, mostly in the northern crop districts. The percentage of state area classified as adequate or surplus in top-soil moisture exceeded 90% throughout the summer until September, except for the week preceding July 9
during which 11% of the state was classified short with the shortages occurring mostly in the western districts. After September, about 25% of the state was classified short in the top soil moisture but fortunately over 90% of the corresponding subsoil moisture was classified as adequate or surplus.

In general, the annual farming cycle for the major crops starts in the spring in the southern parts of the state and migrates northward as temperatures warm. In 1990, about 50% of the tillage operations and fertilizer applications (field preparations) were complete by the first week in April. In the first week of May, about 95% of field preparations were complete and 50% of the corn and 6% of the soybean were planted. Then and after in 1990, the phenological development of the major crops, corn and soybean, was reasonably normal for both crops (Table 1). Overall, 93% of the total state land area was classified as "Land in Farming", with corn, soybean, hay, and oat production areas representing 67% of the total state land area (based on information from G. Miller, personal communication, 1996 and State of Iowa et al., 1990).

2.3. Albedo Data

Composited georegistered albedo data for 1 km² areas covering the State of Iowa over 16 biweekly periods from March 2 to October 11, 1990 were estimated with equally weighted channel-1 and channel-2 counts converted to band reflectance from the United States Geological Survey [USGS] ERDOS Data Center [EDC] 1990 Conterminous AVHRR [for the NOAA-11 Advanced Very High Resolution Radiometer polar orbiting satellite] Data Set [Eidenshink, 1992]. The channel-1 bandwidth was 0.58 to 0.68 μm and the channel-2 bandwidth was 0.725 to 1.10 μm. EDC image analysis was done with the Land Analysis System (LAS) [Ailts et al., 1990]. Georegistration was done with Lambert azimuthal equal area projections based on the USGS Digital Line Graph data with RMSE error less than 1 pixel (the 1 km² areas). Compositing and calibration procedures followed those used in the
construction of the normalized difference vegetation index [NDVI] data, which minimizes cloud effects [Holben, 1986]. Due to edge effects and digitizing, the methodology does not exactly reproduce the area of Iowa previously stated but the error is small [<0.2%]. Albedo data ranged from 0 to 63.5%. The maximum value represents all counts that would have scaled to higher albedo values but, for data processing reasons, were truncated. Generally the maximum value represents clouds, snow, or other non-vegetated bright surfaces.

2.4. Spatial Analyses

Surface albedo is assumed to be a random field; formally, \( \{ \text{albedo}(s), s \in \text{Iowa} \} \) with \( \text{albedo}(s) \) being the surface albedo value at location \( s \). Each of the 16 data sets was examined for large-scale linear and quadratic trends. The latter were generally all significant but very small in magnitude. For each of the 16 data sets two directional semivariograms based on Hawkins and Cressie’s [1984] robust method were estimated along the eastings (537 km maximum) and northings (356 km maximum). This robust semivariogram estimator is given by

\[
\gamma(h) = \frac{1}{2|N(h)|} \sum_{(s_i, s_j) \in N(h)} \left| \text{albedo}(s_i) - \text{albedo}(s_j) \right|^4 / (0.457 + 0.494 / |N(h)|) \tag{1}
\]

where \( |N(h)| \) is the number of distinct pairs in \( N(h) \) with \( N(h) \) given by

\[
N(h) = \{ (s_i, s_j) : s_i - s_j = h \ ; \ i, j = 1, \ldots, n \}, \tag{2}
\]

\( n \) is the number of observations within the region of interest, and \( s \) is the location vector. Eq. (1) represents an estimate of the semivariogram, \( \gamma(h) = \text{Var}(\text{albedo}(s+h) - \text{albedo}(s)) \). Estimation of directional semivariograms was employed so that no assumption of isotropy was needed. The first set of estimated semivariograms was based on the raw data and inherent large-scale variations (trends) were not removed. The second set of estimated semivariograms was based on residuals with large-scale variations removed. The removal was accomplished by a modified two-cycle median polish
patterned after Emerson and Stoto's [1983] method; the modification reduced possible noise effects as well as the grid size for the median polish. The idea of using a median polish to remove large-scale spatial variation was introduced by Cressie [1986]. For each set, a rectangle that inscribed all the data within Iowa was divided into 16 km² blocks. Instead of using all the data, block medians were estimated via the median polish, and then the medians from nonempty blocks were used to estimate residuals for each corresponding block in the original grid. Following the approach of Journal and Huijbregts [1978], directional analyses were done only up to half the possible respective lag distances, 268 lags for the eastings and 176 lags for the northings. Nuggets \(c_0\) for each directional semivariogram were estimated with count-weighted linear least squares regressions on the first three observations in the semivariogram [Cressie, 1991]. When \(c_0 > 0\), \(c_0\) was subtracted from each point in the semivariogram. Thus, for the sets that met this condition, the sill estimator would really be that for the partial sill and further analyses were done on the \(\gamma(h) - c_0\) values.

Following Vanmarcke [1983], a linear scale of fluctuation, \(\theta\), defined as

\[
\theta = 2 \int_0^{\infty} \rho(h) \, dh, \tag{3a}
\]

with

\[
\rho(h) = 1 - \frac{\gamma(h)}{C(0)} \tag{3b}
\]

was calculated for each set and direction. A rectangular rule was used to numerically integrate the correlation function [denoted \(\rho(h)\) in (3a) and (3b) with \(C(0)\) denoting the semivariogram sill in (3b)]. Formally this is

\[
\theta = (1 + 2 \sum_{h=1}^{N} \rho(h)) \Delta h, \tag{4}
\]

where \(\Delta h\) is the increment interval, here 1 km, and \(N\) is the number of increments. An iterative numerical method for estimating \(\theta\) was devised. First the asymptotic variance [the semivariogram sill, denoted \(C_i(0)\)] was
estimated from the median [or a higher percentile if needed] value of \( \hat{y}(h) \). Then \( \rho_1 \) was obtained from
\[
\rho_1(h) = 1 - \frac{\hat{y}(h)}{C_1(0)},
\]
and hence, \( \theta_1 \) was obtained from Eq. (4). At the second iteration \( C_1(0) \) was defined as the median of \( \hat{y}(h) \) for \( \hat{y}(h) > 0 \) and the process was repeated until \( C_1(0) = C(0) \) and \( \theta_1 = \theta \).

An areal scale of fluctuation for each directional pair, \( \alpha \), was estimated as,
\[
\alpha = \theta_x \theta_y,
\]
the product of the northing and easting scales under the assumption of isotropy [a conditional probability definition for \( \alpha \) is given in Vanmarcke [1983], p. 356]. The directional ratio of the two, \( \theta_x/\theta_y \), was used to assess the isotropy assumption. Reported results are \( \theta_x \), \( \theta_y \), \( \theta_x/\theta_y \), and \( \alpha \) for both the raw and detrended data sets along with mean, variance, median, and a robust variance for each of the 16 original albedo data sets. The robust variance estimator used is given by
\[
\sigma^2_\alpha = \left( \frac{1}{N} \right) \left( \frac{1}{n} \sum_{s=1}^{N} | \text{albedo}(s) - \text{albedo}_0 | \right)^2 / (0.457 + 0.494/N)
\]
In Eq. (7) \( \text{albedo}_0 \) was the median of all \( N \) albedo values.

Next, an assessment of bias in the \( \theta \) and resulting \( \alpha \) estimates was conducted based on comparisons. Omnidirectional semivariograms and corresponding radial \( \theta \) estimates were calculated for three selected data sets. The three sets were from the detrended data and were chosen to span the range of the \( \theta_x/\theta_y \) ratio. Specifically, these sets corresponded to the minimum, median, and maximum \( \theta_x/\theta_y \) ratio.

2.5. Temporal Analyses

A first order surface, a plane, was fit to each of the median polished sets as a simple way to model the spatial trend in surface albedo as a function of northing and easting coordinates. The temporal trend over
the season in each of the resulting regression coefficients was then examined with correlation and time series analysis.

The temporal behavior of $\alpha$ was examined with simple autocorrelation analysis on both the original $\alpha$ estimates and the first-difference-in-time $\alpha$ estimates ($\Delta \alpha$, a high pass filter) assuming the time intervals between periods was constant. Though the latter assumption was not strictly true, it was approximately so. Since the data were sparse in time, no time-series modeling or further analyses were done. The Statistical Analysis System (SAS™ v. 6.12) was used to perform all the reported spatial, temporal, and statistical analyses.

3. Results

3.1. Spatial Analysis

As shown in the Fig. 3 boxplots, most of the albedo data were in the range of 10-25% [5th-95th percentiles]. The central tendency and magnitude of variability change throughout the set of observations. Lower values were probably water bodies or very wet soil while higher values were probably clouds since there was very little snow accumulation in the first two periods and none thereafter. Notice in both frames of Fig. 4 the major stream basins are well defined. Throughout the spring and early summer, when much of the land surface was still exposed, the Des Moines Lobe was noticeably visible [Fig. 4 top] but as the crops reached maturity this surface feature disappeared [Fig. 4 bottom]. Of interest, in the Des Moines Lobe, 95% of the total regional area was classified as "Land in Farming" with corn, soybean, hay, and oat production adding up to 83% of the total area [based on information from G. Miller, personal communication, 1996 and State of Iowa et al., 1990]. In the bottom of Fig. 4, crop cover and uniformity of surface-reflective properties were probably

\[\text{Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.}\]
at or near their maximum. Corn normally fully covers the ground after reaching the late vegetative stages [V12-V14 (12 to 14 leaves), G. Benson, personal communication, 1996] which is just prior to the onset of the first reproductive stage, tasseling. In 1990 this phenological period for corn would have occurred mostly in the second half of July. Soybean normally would fully cover the ground in its middle reproductive stages when pod development occurs [R3-R4 (beginning to full pod development), G. Benson, personal communication, 1996]. In 1990, this phenological period for soybean would have occurred mostly in the first half of August. Not surprisingly, the albedo variances were at their lowest during the late summer just after mid August (periods 13-15, Table 1 and Fig. 3) and were at their largest values just prior to this point when rapid changes were occurring either due to plant growth and development (period 7, Table 1 and Fig. 3) or due to the start of harvesting (period 16, Table 1 and Fig. 3).

In Table 1, some of the reported $\theta$-values [Eq. (4)] were estimated based on shortening the length of the integration interval to eliminate noise effects in the tail of the integration interval. When there was a well defined flat sill the $\theta$-estimates converged within a few iterations and agreed within less than one percent, as in Fig. 5, but when the sill was poorly defined, very noisy, curved up or down, the agreement was much less. These problem sets, all but one of which were encountered in the raw data analyses, required shortening of the integration interval. The sets which had no sill or were concave down or continued to grow beyond the first flattened portion may not have valid $\theta$-estimates [Eq. (4)] and resulting $\alpha$-estimates [Eq. (6)]. Moreover there may be a larger-scale phenomenon or trend contamination that current methodology cannot reveal. Thus results from the raw data must be treated with caution. The results from the detrended data are more consistent and probably more credible, perhaps because significant large-scale variations [trend contamination] were removed.
The $\theta$-estimates for the eastings were generally greater than those for the northings. Median values for $\theta_e/\theta_n$ from both the raw data results and the detrended data results were 1.01 and 1.06, respectively. Only in three cases, all from the raw data analyses, were the ratios or their reciprocals more than 1.5. Therefore, on a practical level, these data can probably all be considered isotropic. In assessment of a first-order Markov-process model, Rodríguez-Iturbe [1986] found ratios as high as 2 made little practical difference in the estimation of the correlation area and the resulting integration area. So, the fact that the anisotropy is generally less for the detrended data means that the albedo error structure is probably isotropic. Although the underlying behavior of the albedo data was not characterized in terms of known variance models with presumed underlying properties, as in Rodríguez-Iturbe [1986], the mathematical behavior of the Iowa albedo semivariograms and resulting correlograms is similar. In addition, there is no serious bias in the magnitude of the $\alpha$ estimates. An omnidirectional analysis on the detrended data for period 8, which had $\theta_e/\theta_n=1.01$ [the closest one to the ratio value of one], resulted in an 18% higher value for $\theta$ [based on $(\theta_{\text{radial}} - \sqrt{\theta_e \theta_n})/\sqrt{\theta_e \theta_n}$]. The same analysis on the omnidirectional detrended data for period 16, which had $\theta_e/\theta_n=2.29$ [the highest ratio value], resulted in an 18% lower $\theta$-value while the analysis for period 9, which had $\theta_e/\theta_n=1.12$ [approximately the median $\theta_e/\theta_n$ ratio value], resulted in an 8% higher value. One additional middle value analysis was done. For period 6, which had $\theta_e/\theta_n=1.09$, the result was a 9% higher value. The actual uncertainty in any of the linear or areal estimators is unknown and could require considerable effort to determine. The period 9 comparison suggests that the median $\alpha$-estimate is about 17% low. This margin of error is probably acceptable for most purposes but the end user ultimately needs to be the one who makes the decision. Most likely the impact of a systematic error in the $\alpha$-estimate can be offset by choosing an averaging area larger than the $\alpha$-estimate (for discussion on this point see section 4.1.). If, however, one chooses to do
the radial θ-estimation, the computing time almost doubles on the platform used in this study [from about 6 to over 12 hours per set].

Thus the preceding results between the areal scales of fluctuation and the univariate statistics presented in Table 1 are consistent and reasonable. As might be expected, linear and areal scales were generally large when the overall variance was large and vice versa. In fact, the Pearson correlation coefficient was r = 0.41 [P<0.11] and the Spearman rank correlation coefficient was r = 0.58 [P<0.02]. All θ- and α-estimates from the detrended data were generally smaller than the corresponding ones from the raw data. For the raw data, α varied from <1% to 20% of the area of Iowa and had a median value equal to 3.4%. For the detrended data, the corresponding values were generally <1% with a median value of 0.8%.

3.2. Temporal Analysis

In each and every data set, all the coefficients of the first order surface fit to the spatial trend in the median polished data were very well determined (P<0.0001). The coefficient of determination averaged 40% and ranged from 6 to 67%. Higher order response surfaces or more complicated models could improve the results if a better or further analysis of the spatial trend is needed. The seasonal trend in intercept appears very similar to that of the mean or median and has no significant linear trend in time or first order autocorrelation [Fig. 6]. The easting coordinate parameter also has no significant time trend or first order autocorrelation [Fig. 6]. As shown in Fig. 6, however, the northing coordinate does have a positive linear trend in time [r = 0.69, P<0.003]. In addition the autocorrelation analyses on the parameter trend reveal a positive lag-1 structure (P<0.05) after the original values were put through a high pass filter. This time trend is consistent with and probably due to state patterns in crop development and maturing.

The autocorrelation analyses on α [Eq. (6)] reveal a positive lag-1 structure (P<0.05) for the high-pass-filtered α datasets developed from
each of the original raw and detrended data sets. Since the α-value in the last period for the detrended set was questionable the analysis was redone without the last datum and the positive lag-1 structure was still found. Given that most large area changes in agriculture fields are generally slower than a two-week period, this lag structure makes sense. Analyses on the unfiltered α datasets, however, revealed no such temporal structure.

4. Discussion
4.1. Statistical Considerations

The estimation procedure for θ used in this study was developed in contrast to the more conventional approach of using a parameter estimate obtained after fitting a semivariogram model to the empirical semivariogram. Initially, various well-known weighted least squares fits of semivariogram models, following Gotway's method (1991), were tried but all showed a noticeable lack of fit as well as other problems that would bias the scale-of-fluctuation estimates. In addition, robust definitions of variance were tried in order to estimate the sill but these were generally systematically low. Data in other studies may also be subject to such problems. Possible reasons for a lack of fit in semivariogram models could be that a given model's curvature is too restrictive or that a mixture of variation scales exists requiring a more complex model. A possible reason for the robust variance estimator being below the sill is that Eq. [7] may be more resistant than needed for a given data set.

Rodriguez-Iturbe (1986) and others have advocated the application of random-field theory to hydrology, meteorology, and related sciences. The estimation of an areal scale of fluctuation is needed to determine an averaging area as well as to build spatial and temporal representations for a given process or property. Furthermore, the determination of an averaging area (or time scale) for a random field provides a way to soundly represent an average for the process or property in a domain of interest. In determining an averaging area for each data set, an arbitrary multiple
of each $\alpha$ is chosen. Based on comparing several areal spread models, Rodriguez-Iturbe ([section 7, 1986]) argued from simulations that an averaging area from 3.5 to 4$\alpha$ should be chosen. These simulations introduce variations such as two-to-one anisotropy into the modeled relative variance (the maximum value is scaled to one). The results were plotted as a function of the factor, averaging area/$\alpha$; the functions appeared to converge at factors of four or more. This convergence behavior may be valid for other variance data based on either real data or simulations but the point of convergence could vary somewhat. Figure 7 shows the results of an averaging area analysis using the raw data for period 9. The mathematical behavior of the regression curve is similar to any of the ones developed from several known distributions that were discussed in Rodriguez-Iturbe [1986]. Notice that any averaging greatly reduces the variance. Here, based on the regression model in Fig. 6, the relative change at 4$\alpha$ is 0.014, a very conservative value. Choosing 2$\alpha$ would only double the relative change value. The uncertainty in the albedo variance data probably does not warrant being very conservative because at 2$\alpha$ the 95% confidence interval for the actual relative variance value, 0.54 (the regression prediction is 0.51), is (0.41, 0.75). In terms of the raw albedo data, however, just using a factor of 4 can give huge areas as a portion of the state, especially with the raw data $\alpha$ estimates. For example, in period 9 the portion for the detrended estimate results in 2.9% of the state’s area while the corresponding result from the raw data estimate is 79.8% of Iowa’s area. If the latter result is valid then only modelers that work with GCM size grids can be free of local averaging effects. Many hydrologists are working with much smaller areas.

If the areal scales of fluctuation for Iowa are reasonably reproducible for the same crop-development stage in different years then the results of this work are of consequence across all scales of interest. Consider that in total area Iowa is about 90% of a 2.5° square grid ($1.6 \times 10^5$ km$^2$) or about 14.6 times a 1° square grid ($10^6$ km$^2$). Given the
distributions in Figure 3, for either grid size, global or mesoscale climate modelers would probably happily consider Iowa to be homogeneous in albedo at any given crop-development stage. Very likely, only the ocean surface or a very large desert would be less variable. Moreover, every averaging area determined from the valid $\alpha$ values (estimated from the detrended data) in Table 2 is less than the area of either climate model grid size. Hence climate modelers can soundly employ areal averages for the common grid sizes or reasonable intermediate alternatives, e.g., Iowa Crop Reporting Districts (regular grid) or major land formations (irregular grid). Researchers working on smaller domains are still all right as long as the domain is larger than the largest averaging area. In domains less than the averaging areas suggested by the analysis over the entire state, the analysis should be redone for data only within the domain of interest. Furthermore, as is discussed in section 4.2, researchers working in smaller domains probably have different objectives, ground truth data, and other additional information.

Generalizations for temporal trends and space-time models should probably not be made with data from only one growing season. Climate modeler's can, however, probably use suitable stochastic variations about the annual trend in long-term averages of albedo [Iqbal (1983) lists such data]. Researcher's working at the lower end of the spatial scale probably need more frequent remote-sensing measurements with some ground truth confirmation. Data or estimates from a space-time model may be what is needed.

4.2. Implications for regional albedo and energy balance estimation

In the context of regional evapotranspiration, the areal scales of fluctuation of all input data will need to be considered as a whole. Such assessment can be based on conditional probabilities [Vanmarcke, 1983]. In practice, however, the assessment will be constrained by model choice, model and data-error sensitivities, and the parameter determined to have
the maximum averaging area. For example, the error sensitivity of net radiation to its components (obtainable through remote sensing), such as albedo or emissivity, is well-known and easy to show. Assessing the spatial variability and acceptable error in all of the data inputs for a chosen parameterization is relevant in deciding how the areal averaging should be done. Before generalizing or using the results of this study in actual energy-balance estimation, several primary considerations are apparent that will have to be addressed in future work.

The spatial analysis in this study was only done on one variable - surface albedo, at one unit areal scale - 1 km², for one geographic region - Iowa, and only for one (somewhat atypical) growing season - 1990 [note that the growth and development of the major crops was reasonably normal]. More case studies should be conducted that broaden the environmental representativeness, spatial/temporal scales of the findings, and include more than just surface albedo. Since crop rotation of corn and soybean is a common and widespread practice, do the α estimates and patterns occur in other years? Probably so for Iowa - further analyses are planned to specifically address this question. What about in rangeland or forested regions? Remote-sensing data sets for other places and years are available. Having data sets with multiple variables may provide useful covariate structures which could reduce analysis on further datasets. Other bands and variables can be constructed with the data available in the remote-sensing data sets. Thus it may be that the areal scale of variability for surface emissivity is similar enough to that for surface albedo that they can use the same averaging area or serve as a covariate.

Most importantly, there are other existing or nascent concepts being offered for determining aggregation areas. Wood et al. (1988) proposed the Representative Elementary Area (REA), which is based on minimizing variance as a function of area and has been used in small watershed studies. Flügel (1995) proposed the Hydrological Response Unit (HRU) based on delineating regions by physiographic properties of a drainage basin. These
physiographic properties are to be defined in a geographic information system. Finally, Bloschl and Sivapalan (1995) discussed various proposals of using scales and integrating over different scales based on fractals, or other measures of "self similarity", determined from topographic, other geographic, and geological information. Perhaps some interesting relationships between the areal scale of fluctuation and one or all of the alternatives (REA, HRU, fractal dimension), if available, may become manifest.

All methods, including spatial statistical methodology, require considerable analytical effort, but the alternatives mentioned above require site specific and geographic information along with some arbitrary assessments to characterize the region of interest while areal scales of fluctuation need not require such additional information. Also remote-sensing data are generally available at a comparatively higher spatial density than data used to characterize the region of interest. So the areal scale of fluctuation is probably the best choice for now.

5. Conclusions

In general the surface albedo data sets for each period were not spatially stationary due to large scale deterministic trends which vary in time primarily in the north-south direction. For the nine original raw data sets that seemed to produce reasonable results, the estimated $\alpha$ values ranged from 1.1 to 14.5 times larger than the corresponding $\alpha$ estimates from the detrended albedo data. The median value for this ratio was 2.6. All but the last one of the detrended data sets, however, produced feasible results. For these sets the assumption of isotropy was reasonable. The $\alpha$ values varied throughout the growing season ranging from 340 to 2010 km$^3$, being smallest during the periods corresponding to the final growth stages for the major crops and largest for periods of rapid crop growth and development as well as harvest. Autocorrelation analysis for the $\alpha$ estimates (put through a high pass filter) revealed a positive lag-1 time
structure. Hence ambient conditions affect surface albedo and could be used to selectively sample other albedo data sets. Further case studies encompassing more variables, a broader range of environmental conditions and spatial and temporal scales will refine our understanding of surface albedo. In addition areal scales of fluctuation should be compared with other hydrological concepts suggested for defining aggregation domains.

Acknowledgments. This work was supported by the USDA-ARS National Soil Tilth Laboratory, Dr. J.L. Hatfield, Director. Prof. N.A.C. Cressie, ISU, Ames, IA, provided advice, encouragement, and comments on the manuscript. Dr. P. Doriswamy, USDA-ARS-BARC, Beltsville, MD provided the AVHRR data. Prof. G.A. Miller and Mr. Brian Tiffany, ISU, Ames, IA, provided the crop and land use data. Prof. G.O. Benson, ISU, Ames, IA, provided the crop phenology information. B. Hoyer, IGS, Iowa City, IA, provided the Iowa topographic and landform area information. Mrs. J. Meek graciously edited the work.

References


### Table 1. Phenology for Iowa’s major crops in 1990.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Corn Crop</th>
<th>Dates for</th>
<th>Soybean Crop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Report</td>
<td>50% Completion</td>
<td>95% Completion</td>
</tr>
<tr>
<td>planting</td>
<td>4/22 (1%)</td>
<td>5/05</td>
<td>6/07 (5%)</td>
</tr>
<tr>
<td>emergence</td>
<td>5/06 (3%)</td>
<td>5/24</td>
<td>6/21 (5%)</td>
</tr>
<tr>
<td>first cultivation</td>
<td>6/04 (8%)</td>
<td>6/27</td>
<td>7/24 (12%)</td>
</tr>
<tr>
<td>tassle</td>
<td>7/15 (10%)</td>
<td>7/25</td>
<td>8/05 (10%)</td>
</tr>
<tr>
<td>75% or more silked</td>
<td>7/22 (15%)</td>
<td>8/11</td>
<td>8/13 (15%)</td>
</tr>
<tr>
<td>milk</td>
<td>8/05 (20%)</td>
<td>8/11</td>
<td>8/19 (10%)</td>
</tr>
<tr>
<td>in or past dough</td>
<td>8/16 (10%)</td>
<td>9/17</td>
<td>9/10 (20%)</td>
</tr>
<tr>
<td>mature</td>
<td>9/16 (3%)</td>
<td>10/18</td>
<td>9/16 (5%)</td>
</tr>
<tr>
<td>harvested</td>
<td>9/16 (3%)</td>
<td>10/18</td>
<td>9/16 (5%)</td>
</tr>
</tbody>
</table>

Dates are approximate, they were interpolated from data in the USDC-NOAA/USDA-NASS 1990 reports. The units are month and day. The percentages are land area with original tabulation unit in acres.
TABLE 2. Summary Statistics for the 16 Iowa Surface Albedo Data Sets Recorded in 1990.

<table>
<thead>
<tr>
<th>Period/Dates</th>
<th>No. Obs.</th>
<th>Raw Data</th>
<th>Detrended Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\mu$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\text{km}^2$</td>
</tr>
<tr>
<td>01 Mar. 02 - Mar. 15</td>
<td>145479</td>
<td>17.79</td>
<td>7.35</td>
</tr>
<tr>
<td>02 Mar. 16 - Mar. 25</td>
<td>145479</td>
<td>14.25</td>
<td>5.08</td>
</tr>
<tr>
<td>03 Apr. 01 - Apr. 12</td>
<td>145479</td>
<td>14.96</td>
<td>5.37</td>
</tr>
<tr>
<td>04 Apr. 13 - Apr. 26</td>
<td>145479</td>
<td>16.33</td>
<td>7.94</td>
</tr>
<tr>
<td>05 Apr. 27 - May 10</td>
<td>145479</td>
<td>15.61</td>
<td>7.35</td>
</tr>
<tr>
<td>06 May 10 - May 24</td>
<td>145479</td>
<td>18.10</td>
<td>8.54</td>
</tr>
<tr>
<td>07 May 25 - Jun. 06</td>
<td>145479</td>
<td>18.61</td>
<td>10.82</td>
</tr>
<tr>
<td>08 Jun. 07 - Jun. 21</td>
<td>145479</td>
<td>14.53</td>
<td>4.77</td>
</tr>
<tr>
<td>09 Jun. 22 - Jul. 05</td>
<td>145479</td>
<td>16.43</td>
<td>6.19</td>
</tr>
<tr>
<td>10 Jul. 06 - Jul. 19</td>
<td>145479</td>
<td>16.53</td>
<td>1.87</td>
</tr>
<tr>
<td>11 Jul. 21 - Aug. 03</td>
<td>145479</td>
<td>19.46</td>
<td>4.66</td>
</tr>
<tr>
<td>12 Aug. 03 - Aug. 16</td>
<td>145479</td>
<td>20.01</td>
<td>5.71</td>
</tr>
<tr>
<td>13 Aug. 17 - Aug. 30</td>
<td>145479</td>
<td>16.98</td>
<td>2.60</td>
</tr>
<tr>
<td>14 Aug. 31 - Sep. 13</td>
<td>145479</td>
<td>17.41</td>
<td>2.41</td>
</tr>
<tr>
<td>15 Sep. 14 - Sep. 27</td>
<td>145479</td>
<td>16.58</td>
<td>2.16</td>
</tr>
<tr>
<td>16 Sep. 28 - Oct. 11</td>
<td>145479</td>
<td>15.67</td>
<td>8.55</td>
</tr>
</tbody>
</table>

*Variables: Period, period number for the compositing of data; Remote Sensing Recording Dates, the dates of data collection for the compositing of data; No. Obs., number of observations (1 km$^2$ cells); $\mu$, arithmetic mean of data; $\sigma^2$, corresponding variance about $\mu$; Md, median (50 percentile of data); $\alpha^2$, robust variance (see Eq. [7] in the text); $\theta_e$, linear scale of fluctuation for northings; $\theta_n$, linear scale of fluctuation for eastings; $\theta_e/\theta_n$, ratio of directional $\theta$'s (a simple measure anisotropy); and $\alpha$, areal scale of fluctuation computed from linear $\theta$s assuming isotropy.

†These $\theta$ estimates were recalculated with a reselected integration interval in order to eliminate significant error from the sill data.

‡These $\theta$ estimates are associated with semivariograms that have no sill so they were recalculated with the sill variance set to the maximum observation value.
Figure 1. Landform regions of Iowa based on Prior (1991).
Figure 2. Monthly temperature and precipitation record. The points connected with the lines are the 1990 data, the solid line is for air temperature (left axis), and the dashed line is for cumulative precipitation (right axis). The broken vertical line segments are the deviation from the long-term normal for each climate variable.
Figure 3. Boxplots of the surface albedo data for each of the 16 periods throughout the 1990 growing season in Iowa. Major crop phenological stages and events are marked along the time axis.
Figure 4. Gray scale plots of the surface albedo for periods 2 (March 16-25, top) and 14 (August 31—September 13, bottom). The darker areas absorb more light and the lighter pixels reflect more light.
Figure 5. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_*$. The semivariance data are the solid circles while the correlation data are the open circles. The semivariance data were from the detrended surface albedo data for period 9 [June 22 to July 5].
Figure 6. Time trends in the parameters from first order response surfaces fit to each median polish set.
Figure 7. Model (line) and data (solid circles) for relative variance as a function of averaging area for albedo data in period 9, June 22 - July 5. The base unit for the averaging area, $a$, was 1060 km$^2$. The model for the weighted (by count) least squares regression on the relative variance, $\sigma^2$, is $\sigma^2 = 0.521a^{-0.051}$ with $R^2 = 0.999$. 
GENERAL CONCLUSIONS

A scientifically sound and pervasive methodology for the scaling-up of local data to regional evapotranspiration and energy-balance models is still not manifest. All sorts of problems exist. Three of the primary problems involving data or parameters needed to model surface net radiation are solvable, but widely known, accepted, and practiced scientifically-based norms or standard methods for each given job are not being employed. Solutions developed in this dissertation for these three primary problems offer progress toward the better estimation of regional evapotranspiration.

In any approach, data quality is a crucial facet. Screening procedures for point-source surface data are available but not always employed. Whenever they are, an interactive, visually-based method is often used. Automated data-screening procedures are needed to eliminate human bias and error through consistent identification of unusual observations that exceed climatic or sensor limits or that indicate either an abrupt change in the data stream or a long steady period in it. The rules developed in the first paper that address these problems were also applied to a third set [for Piketon, Ohio] and revealed a siting problem that had not previously been apparent [the station was westward but very close to a highway embankment which diminished sunrise solar radiation values and increased sunset solar radiation values]. Through the consistent identification of unusual observations and events, the automated screening rules can help with data exploration as well as improve data quality.

In all cases implementation of data-checking procedures must be carefully done and recommended guidelines should be carefully followed. Successful implementation and full benefit of such procedures can only come about, however, if conducted as part of combined field operation and data-processing and management quality control protocols. These protocols should address the dissemination of data and system information to a wide range of external users. Feedback from external users would help refine
the process. There are many possibilities for further related research: (1) Rules governing observations not considered in the first paper, e.g., net radiation, could be developed and evaluated. Net radiation could be bound from above by global solar radiation and below by outgoing longwave radiation. (2) A comparison of missing-value algorithms for each data type could prove interesting. Do spatial or temporal linear interpolation work as well as time series? (3) The usefulness and best means of incorporating these point-source screening rules into a system that checks data for a network of stations needs to be considered. For small, local networks, just adding simple paired double-mass analyses may be sufficient. For large dense networks, the rules could be implemented at each station or spatially generalized. (4) The emerging rules for a network of stations are also visual and human-judgement based. What kinds of network rules could be fully automated and generalizable? Formal spatial analysis and expert systems should be considered.

Before the advent of the SAMSON climatic database, daily norms and extremes for many solar radiation and related meteorological variables could not be determined because a 30-year period of record at numerous locations throughout the United States was not available. Consequently, the estimation of clear-day global solar radiation was not climatically based. Moreover, the estimation/confirmation methodology was arbitrary and ad-hoc. With the SAMSON database, clear-day solar radiation at the land surface can now be climatically defined and reasonably estimated for locations throughout the United States when based on broadband-global-solar-radiation models driven by climatically low trends in the turbidity variables. Based on the overall performance of the modified solar radiation models as well as a comparison of the turbidity data to long-term monthly averages at one location, the turbidity data in the SAMSON database are reasonably good for the purpose of estimating maximum possible global solar radiation at the land surface. Furthermore, distinctly different annual trends are evident for normal versus climatic minima data in both
daily aerosol optical depth and daily precipitable water. The latter results may be of interest to air-quality monitoring and regulatory agencies. Some possibilities for further related research follow: (1) The suitability of the broadband models for estimation of subdaily periods needs to be formally considered. (2) Comparison of daily net radiation modeled with some of the available empirical clear day routines to that modeled with the broadband model based on climatic lows in the turbidity variables should be conducted. Is it the same? (3) Tables of the climatic-normal and climatic-low models for both turbidity variables for all SAMSON sites, especially the primary sites, could be helpful to several related branches of atmospheric science. Moreover, they may promote the adoption of the modified broadband-global-solar-radiation models by environmental modelers who need clear-day estimates.

No standard objective method to determine an aggregation domain for spatial data is widely accepted and practiced although one based in random-field theory was suggested over a decade ago. A case study using concepts from random-field theory was conducted on satellite-derived reflectance data. Aerial scales of fluctuation for both stationary and nonstationary assumptions, estimated from remote-sensing surface-albedo data covering the state of Iowa throughout several periods in the 1990 growing season, showed variation throughout the entire time span with surface conditions affecting the estimates. Furthermore, the analysis showed that an averaging area for a typical period should be at least double that of the aerial scale of fluctuation. If the results are reproducible in other years for similar cropping patterns and development stage, then the global, and possibly mesoscale, climate modelers can reasonably assume Iowa and perhaps the whole Corn Belt, is homogeneous in albedo. They could then readily take an average over the whole state, or large subareas, for each developmental crop stage. Smaller areas require more careful consideration of goals and other available information. Further related research is needed and should include the following: (1) Methodology is needed to determine the
uncertainty in linear and areal estimators. (2) Methodology is needed to
more soundly assess the averaging area recommendation. Stochastic
simulation studies may be helpful in evaluating (1) and (2). (3)
Additional data sets and analyses are needed to test the reproducibility of
estimators and stationary assumptions over a broader range of environments,
circumstances, and time. Plans are to test similar Iowa albedo data sets
for other years at selected times in the growing season. (4) Comparisons
are needed to determine relationships [if any] of these estimators with
those from other methods of aggregation. (5) Case studies are needed to
evaluate the means of determining which is the best approach for given
circumstances and goals, especially for studies focusing on the smaller
areas such as for hydrological basins which generally have considerable
additional information available. Perhaps the database for one of the
ongoing large-area field experiments could provide a suitable means of
assessment for some or all of the concepts and further issues mentioned,
not just for the last paper but for any and all of those espoused in this
dissertation.

In summary, the methodology proposed in this dissertation was
formulated on well-known, sound principles and confirmed with real-world
data. It is not intended to be rigid but to serve as a reasonable basis on
which to build. Although need and circumstance will probably require
suitable modification, the adoption of these methods offers standardization
for certain extant practices, promises better estimates of net radiation,
lays groundwork for other methodological development via concrete examples,
and thus can contribute to the advancement of knowledge in the regional
estimation of evapotranspiration and energy balance.
Fig. A1. Clear day solar radiation model results for Ames, IA, model 1 in (a) and model 2 in (b). In each figure the top solid line is for the extraterrestrial model, the upper broken line is for the LLI inputs, the lower solid line is for the climatic low inputs, and the bottom broken line is for the climatic normal inputs. The solid circles are the selected clear day data.
Fig. A2. Clear day solar radiation model results for Bismarck, ND, model 1 in (a) and model 2 in (b). In each figure the top solid line is for the extraterrestrial model, the upper broken line is for the LLI inputs, the lower solid line is for the climatic low inputs, and the bottom broken line is for the climatic normal inputs. The solid circles are the selected clear day data.
Fig. A3. Clear day solar radiation model results for Dodge City, KS, model 1 in (a) and model 2 in (b). In each figure the top solid line is for the extraterrestrial model, the upper broken line is for the LLI inputs, the lower solid line is for the climatic low inputs, and the bottom broken line is for the climatic normal inputs. The solid circles are the selected clear day data.
Fig. A4. Clear day solar radiation model results for Wooster, OH, model 1 in (a) and model 2 in (b). In each figure the top solid line is for the extraterrestrial model, the upper broken line is for the LL1 inputs, the lower solid line is for the climatic low inputs, and the bottom broken line is for the climatic normal inputs. The solid circles are the selected clear day data.
Fig. B1. Plot (a) is for the raw surface albedo data from period 1 [March 2 - 15]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance (γ, left axis) and correlation (ρ, right axis) functions for estimating the linear scale of fluctuation, θ. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B2. Plot (a) is for the raw surface albedo data from period 2 [March 16 - 25]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_s$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B3. Plot (a) is for the raw surface albedo data from period 3 [April 1 - 12]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\phi$, right axis) functions for estimating the linear scale of fluctuation, $\theta_x$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B4. Plot (a) is for the raw surface albedo data from period 4 [April 13 - 26]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_s$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B5. Plot (a) is for the raw surface albedo data from period 5 [April 27 - May 10]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\ell$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B6. Plot (a) is for the raw surface albedo data from period 6 [May 10 - 24]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_2$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B7. Plot (a) is for the raw surface albedo data from period 7 [May 25 - June 6]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_s$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B8. Plot (a) is for the raw surface albedo data from period 8 (June 7 - 21). Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\xi$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B9. Plot (a) is for the raw surface albedo data from period 9 [June 22 - July 5]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B10. Plot (a) is for the raw surface albedo data from period 10 [July 6 - 19]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($p$, right axis) functions for estimating the linear scale of fluctuation, $6_q$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B11. Plot (a) is for the raw surface albedo data from period 11 [Jul. 21 - Aug. 3]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_0$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B12. Plot (a) is for the raw surface albedo data from period 12 [Aug. 3 - Aug. 16]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($p$, right axis) functions for estimating the linear scale of fluctuation, $\beta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B13. Plot (a) is for the raw surface albedo data from period 13 [Aug. 17 - 30]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_s$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B14. Plot (a) is for the raw surface albedo data from period 14 [Aug. 31 – Sep. 13]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance (γ, left axis) and correlation (ρ, right axis) functions for estimating the linear scale of fluctuation, $S_e$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B15. Plot (a) is for the raw surface albedo data from period 15 [Sep. 14 - 27]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance (γ, left axis) and correlation (ρ, right axis) functions for estimating the linear scale of fluctuation, θx. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B16. Plot (a) is for the raw surface albedo data from period 16 [Sep. 28 - Oct. 11]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\ell$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B17. Plot (a) is for the raw surface albedo data from period 1 [March 2 - 15]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\delta_0$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B18. Plot (a) is for the raw surface albedo data from period 2 [March 16 - 25]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\xi_y$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B19. Plot (a) is for the raw surface albedo data from period 3 [Apr. 1 - 12]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\Theta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B20. Plot (a) is for the raw surface albedo data from period 4 [Apr. 13 - 26]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\xi_0$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B21. Plot (a) is for the raw surface albedo data from period 5 [Apr. 27 - May 10]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($p$, right axis) functions for estimating the linear scale of fluctuation, $\delta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B22. Plot (a) is for the raw surface albedo data from period 6 [May 10 - 24]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $S_\theta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B23. Plot (a) is for the raw surface albedo data from period 7 [May 25 - Jun. 6]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_x$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B24. Plot (a) is for the raw surface albedo data from period 8 [Jun. 7 - 21]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B25. Plot (a) is for the raw surface albedo data from period 9 [Jun. 22 - Jul. 5]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance (γ, left axis) and correlation (ρ, right axis) functions for estimating the linear scale of fluctuation, 64. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B26. Plot (a) is for the raw surface albedo data from period 10 (Jul. 6 - 19). Plot (b) is for the corresponding detrended surface albedo data. Semivariance (γ, left axis) and correlation (ρ, right axis) functions for estimating the linear scale of fluctuation, $S_v$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B27. Plot (a) is for the raw surface albedo data from period 11 [Jul. 21 - Aug. 3]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_\gamma$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B28. Plot (a) is for the raw surface albedo data from period 12 (Aug. 3 - 16). Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($y$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_{0}$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B29. Plot (a) is for the raw surface albedo data from period 13 [Aug. 17 - 30]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B30. Plot (a) is for the raw surface albedo data from period 14 [Aug. 31 - Sep. 13]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\theta_x$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B31. Plot (a) is for the raw surface albedo data from period 15 [Sep. 14 - 27]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\delta_q$. The semivariance data are the solid circles while the correlation data are the open circles.
Fig. B32. Plot (a) is for the raw surface albedo data from period 16 [Sep. 28 - Oct. 11]. Plot (b) is for the corresponding detrended surface albedo data. Semivariance ($\gamma$, left axis) and correlation ($\rho$, right axis) functions for estimating the linear scale of fluctuation, $\xi$. The semivariance data are the solid circles while the correlation data are the open circles.


Hubbard, K.G., 1988: Collection quality control and dissemination of weather data for irrigation and other operations in an automated setting.


ACKNOWLEDGMENTS

I sincerely thank the many people and institutions whose guidance and assistance made the completion of this dissertation possible. This work was supported by the USDA-ARS-MWA National Soil Tilth Laboratory (NSTL), Dr. J.L. Hatfield, Laboratory Director and my major professor. Prof. R.E. Carlson, Dstg. Prof. N.A.C. Cressie, Assoc. Prof. W.W. Simpkins, and Prof. D.N. Yarger from Iowa State University and Dr. D.B. Jaynes from the NSTL served on my committee. Dr. J. Prueger, Mr. W. Oesterreich, Mr. S. Farris, and various other members of the NSTL staff have provided data, help, or other services as needed. Dr. R.L. Pfeiffer also from the NSTL has often provided a laugh when I needed one.

I dedicate this dissertation to my wife, Jean. Her love, support, and guidance have been vital.
NAME OF AUTHOR: David Winter Meek

DATE AND PLACE OF BIRTH: March 2, 1952, Richmond, Indiana

DEGREES AWARDED:
B.A. in Physics & Mathematics, University of California, Santa Cruz, 1977
M.S. in Applied Mathematics, Rensselaer Polytechnic Institute, 1979

HONORS AND AWARDS:
Member, ΠΣΑ (1994)
Member, ΦΦ (1995)
Iowa State Univ. PACE Scholarship (1991, declined)
Rensselaer Departmental Scholarship (1978 and 1979)
San Francisco City College Dean’s List (Fall 1973, Spring 1974) and

PROFESSIONAL EXPERIENCE:
1978, Student Associate-Graduate, I.B.M.-G.P.D., San Jose, CA
1979-1990, Mathematician, USDA-ARS-PWA Water Mgt. Res. Unit, Fresno, CA
1990-present, Mathematician, USDA-ARS-MWA Natl. Soil Tilth Lab., Ames, IA

SELECTED PUBLICATIONS:
Ayars, J.E. and D.W. Meek. 1994. Drainage load-flow relationships in arid
Eng. 117(6): 975-977.
Heteroscedasticity in whole plant growth curves developed from
Banuelos, G.S. and D.W. Meek. 1990. Accumulation of selenium in plants
Computational approach to assess actual transpiration from aerodynamic
Bar Yosef, B. and D. Meek. 1987. Selenium sorption by kaolinite and
1984. A generalized relationship between photosynthetically active
1984. Automated weather data collection for research on irrigation