

ICARDA Agro-Climate Tool

Technical Description

Steve Mauget¹ and Eddy De Pauw²

¹U.S. Department of Agriculture-Agricultural Research Service
USDA Plant Stress and Water Conservation Laboratory, Lubbock, Texas
3810 4th Street, Lubbock, Texas, 79415
1-806-723-5237
Steve.Mauget@.ars.usda.gov

International Center for Agricultural Research in the Dry Areas
P.O. Box 5466, Aleppo,
Syrian Arab Republic
E.DE-PAUW@CGIAR.ORG

Abstract

A Visual Basic agro-climate application developed by climatologists at the International Center for Agricultural Research in the Dry Areas and the U.S. Department of Agriculture is described here. The application's climate database consists of weather generator parameters derived from the station data of 649 meteorological stations. From those parameters the program calculates climate statistics over arbitrarily defined periods within summer or winter growing seasons at user-selected latitude-longitude coordinates. The statistics reported include: crop evapotranspiration estimates derived from the FAO-56 single crop coefficient algorithm, probabilities of exceedance of both cumulative rainfall and growing degree days, the probability that minimum and maximum daily temperatures will exceed user-defined temperature thresholds, and the probability of heat stress, cold stress and dry periods of varying duration.

1. Introduction

The mandate area of the International Center for Agricultural Research in the Dry Areas (ICARDA) extends from northwest Africa to central Asia and is home to more than 755 million people (Fig. 1). The growing regions of those countries are climatically diverse (De Pauw, 2000; Ryan et al. 2006; De Pauw, 2007), but very little information about climate and abiotic plant stress is available to ICARDA plant breeders, agronomists, and hydrologists. The ICARDA Agro-Climate Tool (hereafter, the “application”), a Visual Basic application that can be run on Windows 2000, XP, and Vista operating systems, was developed to address that need. A schematic screenshot of the application can be found in Fig. 2.

Under ideal circumstances the ICARDA mandate countries would each possess dense meteorological networks, and each station in those networks would provide unbroken daily weather records over a recent 30 year period. With those conditions an agro-climate application similar to that described here could be based on observed weather data similar to the daily temperature and precipitation measurements provided by the United States cooperative (National Weather Service, 2000) and historical climatological networks (Easterling et al., 1999). However, actual data availability over the ICARDA countries is far from ideal. Although station density is sufficient over many agricultural areas, many of those stations have considerable amounts of missing data. Moreover, requiring data records of 30 year duration would leave many areas with no or insufficient station coverage. Thus one of the leading challenges in the application’s development was to estimate seasonal climate variability from shorter, and sometimes fragmentary, data records. This challenge was addressed here by generating primary daily weather variables from modified GEM6 (Hanson et al., 1994) weather generator code, based on the monthly statistics of the available daily data. Secondary variables (daily dew point temperature, short-wave surface radiation, net outgoing long-wave radiation and reference grass evapotranspiration) were derived from primary variables using algorithms drawn from the FAO’s ‘Guidelines for Computing Crop Water Requirements’ (Allen et al., 1998), hereafter referred to as ‘FAO-56’. Daily crop evapotranspiration (ET) values were then derived from the reference grass ET values using the FAO-56 single crop coefficient method.

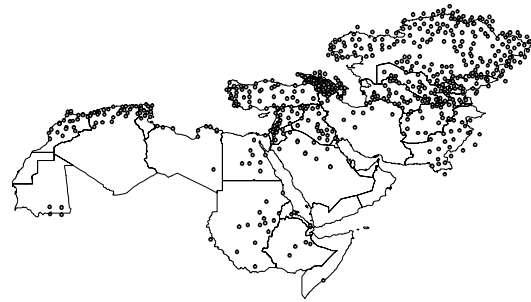


Fig. 1. The ICARDA mandate region, with locations of the 649 meteorological stations used to provide data for the ICARDA Agro-Climate Tool.

2. Data

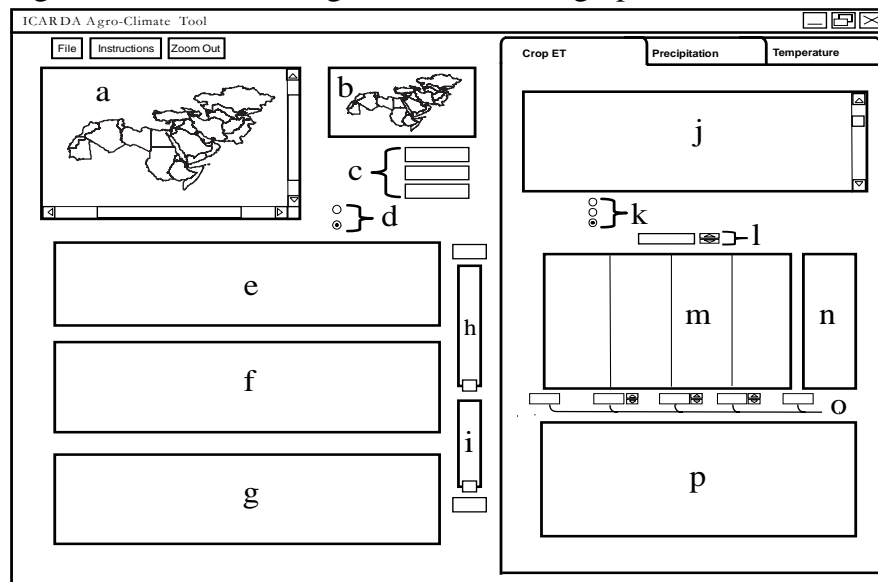
The application’s climate statistics are derived from two data sets that provide daily records of minimum and maximum temperature and precipitation. The main data source is the Global Daily Summary Data (GLDS) set (National Climatic Data Center, 1994), which provides data for ICARDA growing regions at 590 meteorological station locations. The period of record for the GLDS data set is October 1977 to December 1991. The secondary data source is the Global Daily Climatology Network data (GDCN; National Climatic Data Center, 2002), which provides records of primary daily variables at 59 additional locations. Data from GDCN stations is of varying duration, but in some cases begins in the early 20th century. However, because the application’s operation involves the averaging of weather generator parameters from different stations to derive parameters for locations between stations, those parameters must be derived from data over a uniformly defined period. As a result, the decision was made to limit the calculation of statistics from GDCN data to the GLDS data period, i.e., 1977-1991.

3. Statistics Calculation

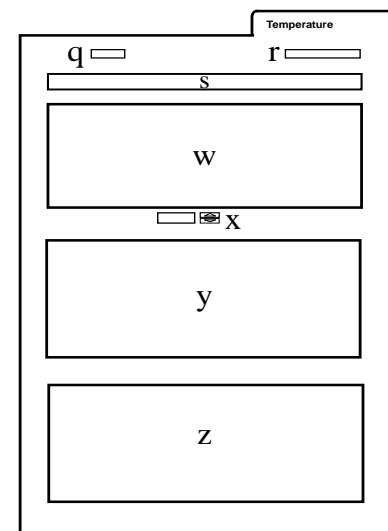
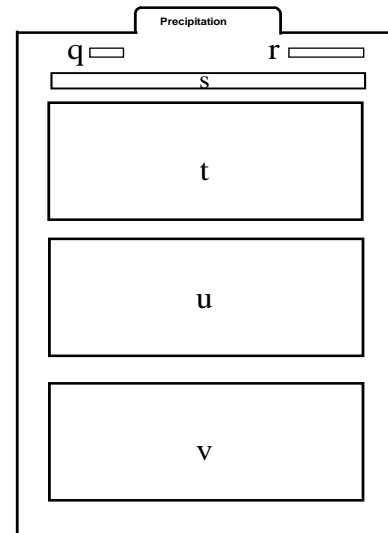
3.1 Data Sampling Requirements

The lack of long-term daily station data, and the intermittent nature of much of the meteorological data that was available, was a limiting factor in the calculation of the application’s climate statistics. These problems were addressed through modifications to the original GEM6 code of Hanson et al. (1994) and by imposing minimum data sampling requirements. An additional strategy for addressing this underlying data problem requires the end-user to use the application in a way that acknowledges the possibility of the resulting sampling uncertainties.

Fig. 2. The ICARDA Agro-Climate Tool graphical user interface.



- a) Location selection map.
- b) Pan view map.
- c) Longitude, latitude, elevation of selected location.
- d) Growing season display selector.
- e) Annual cycle of min. and max. temperature (T) for selected location.
- f) Probability of min. and max. T exceeding user defined heat stress and cold stress thresholds, and probability of rainfall.
- g) Annual cycle of mean rainfall on rainy days.
- h) Heat stress threshold (HST) slider control and display.
- i) Cold stress threshold (CST) slider control and display.
- j) Crop selection scroll list.
- k) Soil type selector.
- l) 2 meter wind speed spinner selector and display.
- m) Crop ET, precipitation, and irrigation demand distributions for four crop growth periods.
- n) Crop ET, precipitation, and irrigation demand distributions for entire growing season.
- o) Crop growth period date displays and spinner selectors.
- p) Probability of exceedance curve for Crop ET, precipitation, or irrigation distribution selected in (n) or (o).
- q) Day of Year cursor indicator for the (s) period selector control.
- r) Period selected by the (s) period selector control.
- s) Period selector control for precipitation or temperature tab.
- t) Probability of exceedance curve for cumulative precipitation for period indicated in (r) on precipitation tab.
- u) Probability that daily rainfall amounts will fall in one of five categories (< 5mm, 5-10 mm, 11-15 mm, 16-20 mm, 21-25 mm, > 25 mm).
- u) Probability of dry periods of varying lengths (< 3 days, 3-5 days, 6-8 days, 9-11 days, 12-14 days, > 14 days).
- w) Probability of exceedance curve for growing degree days (GDD) for period indicated in (r) on temperature tab.
- x) GDD temperature threshold spinner selector and display.
- y) Probability of heat stress periods of varying duration (< 3 days, 3-5 days, 6-8 days, 9-11 days, 12-14 days, > 14 days), for heat stress threshold defined by (h).
- z) Probability of cold stress periods of varying duration (< 3 days, 3-5 days, 6-8 days, 9-11 days, 12-14 days, > 14 days), for cold stress threshold defined by (i).



Normally, GEM6 code calculates weather statistics over 24 bi-weekly periods of the year. However, the limited daily data availability over many agriculture regions, and the data sampling requirements imposed here, made bi-weekly averaging impractical. As a result, GEM6 weather generator parameters were derived from monthly statistics of precipitation, and minimum and maximum temperature. A complete daily data record during October 1977–December 1991 would result in approximately $14 \times 30 = 420$ daily weather measurements contributing to each month's precipitation and temperature statistics. But in ICARDA agricultural areas outside of the former Soviet Union gaps in daily weather records were frequent. To provide adequate station coverage over those areas while also calculating reasonably representative monthly statistics, a minimum sampling threshold of 60 days for monthly statistics was imposed. Thus for example, the mean maximum temperature for January at a station location might be based on as few as 60 daily maximum temperature measurements. Deriving an average from such a limited number of measurements can lead to sampling error in the resulting statistic. The magnitude of sampling error is proportional to $N^{-1/2}$, where N is the number of measurements (Mendenhall et al., 1990). A sample mean calculated from $N=60$ measurements can lead to an error (i.e., the difference between the true, population mean and the mean calculated from a 60 day sample) as large as $0.26 * \sigma$, where σ is the standard deviation of the population distribution. Errors for monthly rainfall transition probabilities (i.e., p00, p10 in Section 3.2), and daily temperature cross correlation values used in the GEM6 multivariate temperature generation scheme (i.e., Eqs 12-15 in Hanson et al., 1994) may be of similar magnitude.

3.2 Calculation of Monthly Statistics

The lack of continuous daily data over ICARDA agricultural regions also influenced the choice of statistics that were calculated from the data. The original GEM6 code calculates two sets of temperature statistics: the mean and standard deviation of maximum and minimum temperature during days with rain, and identical statistics for days without rainfall. But dividing the temperature data by rainfall condition would have caused many stations to fail the minimum monthly sampling requirement described above, producing a sparse meteorological network. As a result, temperature statistics here were calculated over all days, both wet and dry. Thus for each station, temperature variation throughout the

year was described through one set of statistics that describe the mean and standard deviation of daily minimum and maximum temperature for each month of the year.

The GEM6 weather generator uses multivariate regressive and autoregressive relationships to derive daily anomalies of maximum and minimum temperature (t_{max} , t_{min}) and short-wave surface radiation (s_{rad}) based on the current and previous day's anomaly values (Eqs. 12-15 Hanson et al. (1994)). The 3×3 'A' and 'B' matrices defining these regressive relationships are derived from cross correlation values calculated between t_{min} , t_{max} , and s_{rad} at 0 and 1 days lag. But because daily s_{rad} values are not available over the ICARDA mandate region, temperature generation here is based on 2×2 matrices that are derived from cross correlation values calculated only between daily t_{min} and t_{max} values. Daily shortwave surface radiation values were estimated via Hargreaves relationship (Section 6). Although GEM6 calculates 'A' and 'B' matrices for each month, the limited availability of daily data at many station locations made the calculation of daily correlation and cross-correlations on a monthly basis impractical. As a result, the application calculates only one 'A' and 'B' matrix per station, which are in turn derived from annual averages of t_{min} and t_{max} correlation and cross correlation values.

The probability that a day will be wet in the GEM6 generation scheme depends on two sets of monthly statistics that describe the probability that a dry day will be followed by a dry day (p00) during each month, and the probability that a wet day will be followed by a dry day (p10) during each month. GEM6 code normally assigns the amount of rain that falls on a wet day using a mixed exponential distribution. However, here it was found that a three-parameter mixed exponential distribution did not perform noticeably better than a simple one-parameter exponential distribution. As the parameter for an exponential distribution is the expectation of the distribution's variable, (Mendenhall (1990)), the exponential parameter used here is the average of the daily rainfall totals, calculated for each month of the year.

4. GEM6 generation of primary synthetic variables.

A flow chart tracing the generation of synthetic daily records of primary daily meteorological variables, i.e., t_{min} , t_{max} and precipitation, can be found in Fig. 3.

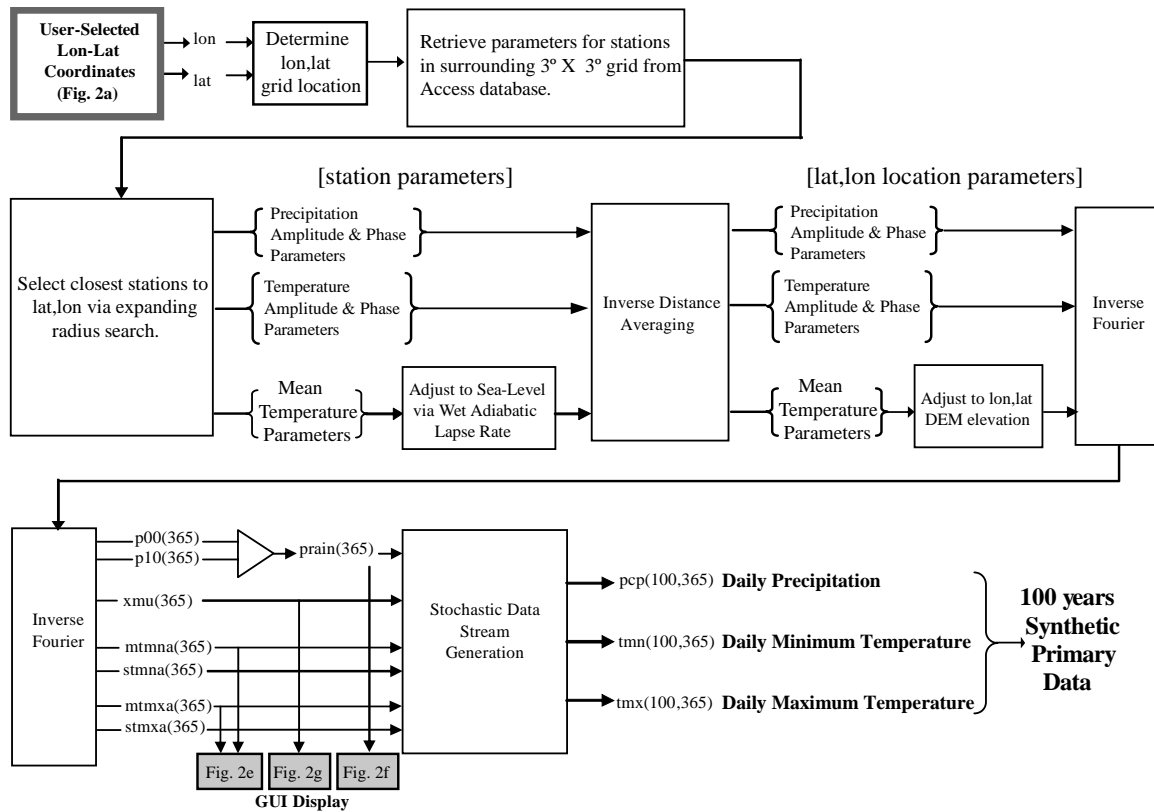


Fig. 3. Flow chart of the generation of synthetic records of daily minimum and maximum temperature and rainfall at a user-selected latitude and longitude. Gray outlined elements indicate a user control, and gray shaded elements indicate a display marked in Fig. 2.

4.1. GEM6 Fourier Parameter Calculation and Storage.

The annual cycles of monthly temperature and precipitation statistics described in Section 3 are interpolated to daily variability in the application by solving for the first three annual Fourier harmonics of the monthly statistics, and then using those harmonics to reconstruct a smoothed version of the annual cycle through an inverse transform. The results of the Fourier transform are stored here in an Access™ database, which provides the inputs for the application's daily weather generation scheme. In addition the database stores information about each station's longitude, latitude, and elevation, and the elements of the 'A' and 'B' correlation and cross-correlation matrices.

4.2 Spatial Interpolation of GEM6 Parameters Between Stations

The VB code calculates a user-selected location's GEM6 parameters as an inverse-distance weighted average of the Fourier parameter sets and the 'A' and 'B' matrix elements of neighboring meteorological stations.

Sets of nearest neighbor stations for ICARDA agricultural areas - the yellow shaded areas in the large locator map (Fig. 2a) - are defined by a second Access database. That database table divides the yellow area into 792 1° longitude by 1° latitude grid areas. The neighboring stations for a 1° by 1° grid area are the stations that lie within a 3° longitude by 3° latitude grid that surrounds that central 1° by 1° grid.

Once the VB code determines which 1° by 1° grid contains the selected location, the Fourier parameter sets for the grid's neighboring stations are then retrieved from the application's primary GEM6 parameter database. The distances between the selected location and those stations are then calculated, and the stations are then sorted according to their distance from the selected location. If the nearest station is within 20 kilometers, then that station's parameters are assigned to the location. Otherwise, the location's parameters are calculated as a distance weighted average of the nearest neighboring stations using an expanding radius search

algorithm. In some areas where station coverage is sparse (e.g., Sudan and Ethiopia) this algorithm could cause a location's parameters to, in the worst case, be averaged from the parameters of stations ~ 200 km away.

4.3 Maximum and Minimum Daily Temperature Generation

To account for the effects of elevation on a selected location's interpolated temperature variation, the Fourier parameters for the mean temperature of the selected neighboring stations are adjusted to sea-level before inverse-distance averaging. This adjustment assumes a mean wet adiabatic atmospheric lapse rate of -6.5°C/km . The adjusted mean temperature parameters, and all the remaining amplitude and phase angle Fourier parameters for all the selected surrounding stations are then averaged using an inverse distance² averaging scheme. After all of the selected location's Fourier parameters have been estimated in this way, the location's mean maximum and minimum temperature parameters at sea level are then adjusted – in most cases decreased, as most stations are above sea-level – to the location's elevation as defined by the GTOPO30 digital elevation model (USGS, 2006). This second elevation adjustment also assumes a wet adiabatic atmospheric lapse rate.

After the entire Fourier parameter set for the selected location has been spatially interpolated, the location's annual cycles of the mean and standard deviations of daily maximum temperature and of the mean and standard deviations of daily minimum temperature are then constructed through an inverse Fourier transform. The annual cycles of mean daily maximum and minimum temperature are used to describe the location's annual temperature cycle in the top graph of the Three-Pane display on the application's left side (Fig. 2e). The four annual mean and standard deviation annual cycles, and the location's spatially interpolated 'A' and 'B' matrices, are then used to generate stochastic streams of daily maximum and minimum temperature. These streams are stored as 100 years of synthetic temperature variation in two 100 X 365 arrays.

4.4 Daily Precipitation Generation

A selected location's precipitation Fourier parameters are estimated using the same inverse distance weighting method used to interpolate temperature parameters.. The resulting parameters for the location are then inverse transformed into three annual cycles: the annual cycle of p00, the annual cycle of p10, and the annual cycle of the exponential rainfall

distribution parameter (XMU). The annual cycle of the probability that rain falls on a given day of the year is derived from the p00 and p10 probabilities via Eq. 4 of Hanson et al. (1999)

$$P_w(\text{day } n \text{ is wet}) = \frac{1 - p00(n)}{1 + p10(n) - p00(n)}, n = 1, 365 \quad (1)$$

These probabilities for each day of the year are used to graph the location's annual cycle for daily rainfall probability in the middle graph of the Three-Pane display on the application's left side (Fig. 2f). Because the exponential rainfall distribution parameter is equal to the average rainfall amount on wet days, the XMU annual cycle is used to graph the average rainfall amount on wet days in the bottom graph of the Three-Pane display (Fig. 2g). The annual cycles of XMU and the p00 and p10 daily rainfall probabilities are used to generate 100 years of synthetic precipitation data which is stored in a 100 X 365 array.

5. Comparison of real vs. generated primary variable statistics.

The application estimates the statistics of real daily temperature and precipitation data, and also statistics of secondary daily variables derived from those primary variables. Because weather generators can produce continuous streams of artificial weather data from incomplete data records, daily weather variation is generated here from weather generator parameters derived from real data. As a result, the accuracy of the application's reported statistics depends on the equivalence of real data statistics with those of the corresponding generated weather data streams.

That accuracy is checked here using data from a network of four central U.S. meteorological stations consisting of Carrizozo, New Mexico, Ballinger and Clarksville in Texas, and Pierre, South Dakota (Fig. 4). These stations are part of the U.S. Historical Climatology Network (USHCN) and provide almost complete records of daily minimum and maximum temperature and precipitation during 1976-2005. During that period each station has no more than 3% missing data in any of the three primary meteorological variables. Days with missing data were filled in with artificial values generated by Hansen et al's (1994) unmodified GEM6 configuration.



Fig. 4. Locations of the USHCN meteorological stations used for verifying GEM6 weather generator output.

Statistics derived from real data were compared with the corresponding statistics of synthetic data generated from the application's modified GEM6 code under three conditions. In the first condition, the application's monthly weather generator statistics were calculated with real daily data from each year of 1976-2005. These synthetic results will be referred to as "Ideal" (I), as climate statistics are ideally calculated from 30 years of complete data. In the second condition, monthly statistics were calculated with data from 1978-1991. These results will be referred to as "Best" (B), as 14 years of continuous data reflects the best possible data conditions in the ICARDA station data. In the final condition, each month's statistics were calculated from 9 randomly selected 7 day sequences of data within that month during 1978-1991. These results are referred to as "Worst" (W), as they correspond closely to the worst case 60 day sampling threshold conditions imposed here on the ICARDA station data.

5.1 Annual Temperature and Precipitation Climatology

The application's Three-Pane display (Figs. 2 e,f,g) shows the annual cycle of a selected location's mean minimum and maximum temperature (T_{\min} , T_{\max}), precipitation probability (P_p) and mean precipitation amount (P_a), as represented by each variable's inverse Fourier transform. The Fig. 5 scatterplots compare the yearly cycles found in the USHCN station's real data with cycles of corresponding generated data under the three monthly statistics conditions. These test's annual cycles are represented by statistics calculated over the year's 26 bi-weekly periods. Those statistics are calculated from the four station's 30 years of real data and also from 100 years of data generated under the I, B, and

W conditions. That figure's twelve scatterplots compare the bi-weekly statistics of observed P_p , P_a , T_{\min} , and T_{\max} , on the X axes, with the same two week period's generated statistic on the Y axes. The first column's three plots (Figs. 5a-c) compare real P_p with generated P_p based on the I, B and W monthly statistics condition. The second column (Figs. 5d-f) consists of similar plots comparing the bi-weekly percentage of wet days. The third and fourth column plots (Figs. 5g-l) compare averages of minimum and maximum temperature. Each scatterplot combines the 26 scatterpoints for all four stations on one plot, thus each plot contains 104 points. The aggregate rms error of the 104 scatterpoints are also shown on each plot.

In Fig. 5 the two precipitation variable's scatter is noticeably greater than that of the temperature variables under each of the three data conditions. In the six T_{\min} , and T_{\max} scatterplots the rms error varies between 1.32–1.42 °C, while the range of variability in the minimum and maximum temperature cycles is ~30.0 °C. By contrast, the 5.9% rms error of the Real vs. Worst case comparison for P_p (Fig. 5c) is about one fourth of the ~25% annual range in precipitation probability. In the precipitation variables the increase in mean error is particularly clear as the amount of data used to calculate the weather generator's monthly statistics decreases. As data conditions go from Best to Worst case conditions the rms error in the P_p annual cycles almost doubles in Figs. 5b and c. Similarly, although the Ideal (5.58 mm) and Best case (6.24 mm) mean error in precipitation amount in Figs. 5d and e are comparable, that error increases to 10.43 mm under Worst case data conditions (Fig. 5f). The weather generator's greater error in reproducing precipitation's annual cycle, even under Ideal data conditions, might be traced to the fact that, when compared to temperature, precipitation's annual cycle calculated over bi-weekly periods is relatively noisy. In the GEM6 rainfall generation scheme a day's wet or dry condition is defined by comparing random number generation with Eq. 4's p_{00} and p_{10} variables. Because those variables are interpolated here to daily resolution via an inverse Fourier transform of the first three annual harmonics, they are highly smoothed. This smoothing, combined with the noisy nature of real precipitation probabilities, may lead to random errors in the rate of rainfall generation over the annual cycle. Although the cycles of Fourier-smoothed P_p and P_a are displayed as green traces in the Three-Pane Display (Figs. 2f,

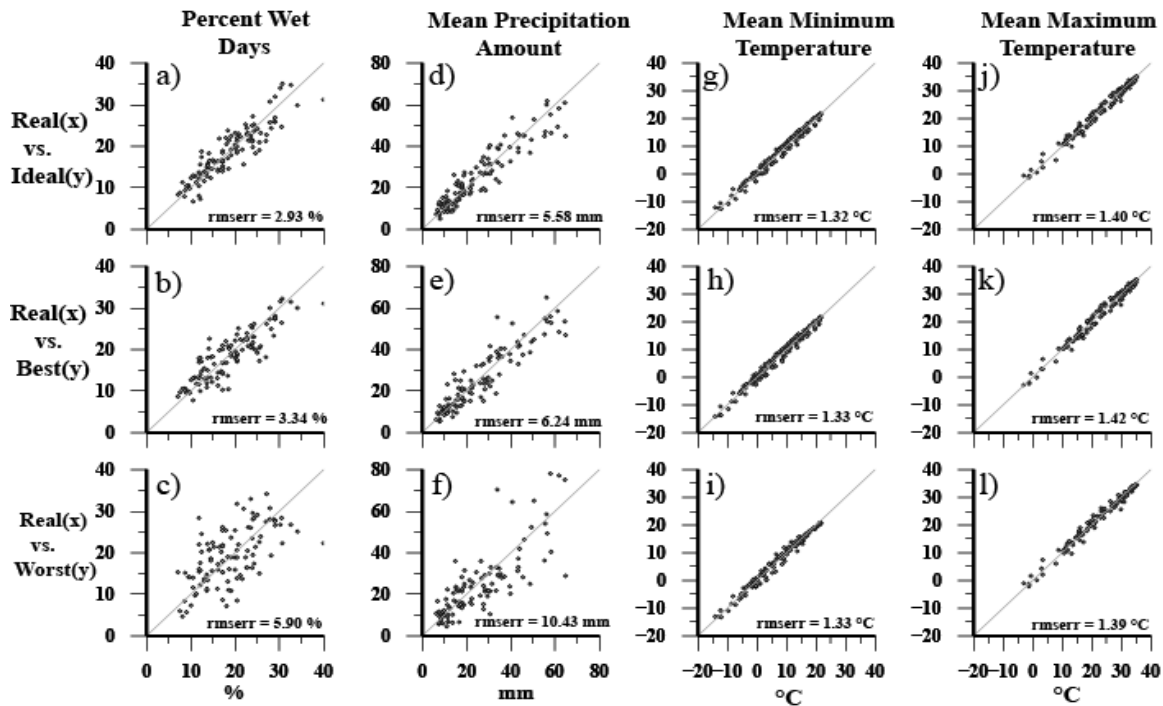


Fig. 5. a-c) Scatterplots of real vs. generated bi-weekly percent wet days throughout the annual cycle for the 4 USHCN stations in Fig. 4, based on the I(a), B(b) and W(c) monthly statistics condition. e-f) As in (a-c) for mean bi-weekly precipitation amount. g-i) As in (a-c) for mean bi-weekly minimum temperature. j-l) As in (a-c) for mean bi-weekly maximum temperature. Root mean square error figures are the aggregate rms error of each plot's 104 scatterpoints.

and g), they are meant as qualitative representations of the real, and much noisier annual precipitation climatology. But the relative agreement between real and generated temperature statistics in Figs. 5g-l shows that, even under the Worst case data conditions, the Three-Pane display's red and blue Fourier-smoothed T_{\min} , and T_{\max} cycles (Fig. 2a) more directly reflect real annual temperature variation.

5.2 Seasonal Precipitation Climatology

Figure 5 compares bi-weekly means, but the application can present climate information calculated over seasonal periods with user-selected beginning and end dates. On the Precipitation Tab the probability of exceedance of total seasonal precipitation is displayed (Fig. 2t), as are bar charts showing the probability of daily rainfall amounts (Fig. 2u) and the probability of dry spells of various duration (Fig. 2v).

Figure 6a's bar and whisker diagrams show distributions of total summer (May 15–Oct. 15)

precipitation in the real data and in the generated data under the Ideal, Best and Worst case data conditions. Those diagrams are shown for each of the four USHCN stations, and divide the distributions into equal 33rd percentiles (terciles). Terciles were defined from 30 years of summer precipitation in the real data, and from 100 summers of synthetic data generated under each data condition. Generally, the distributions produced by the weather generator under each condition tracks the real data's station-to-station variability. For example, the observed west-to-east increase in summer rainfall between Carrizozo and Clarksville is also found in each of the generated distributions. But at Clarksville the central bars of those distributions show a positive bias relative to the real distribution as data conditions go from Ideal to Worst, while at Pierre there is an opposite tendency. But overall, the GEM6-generated distributions at the four sites reproduce the observed range of 1976-2005 summer precipitation fairly well. This suggests

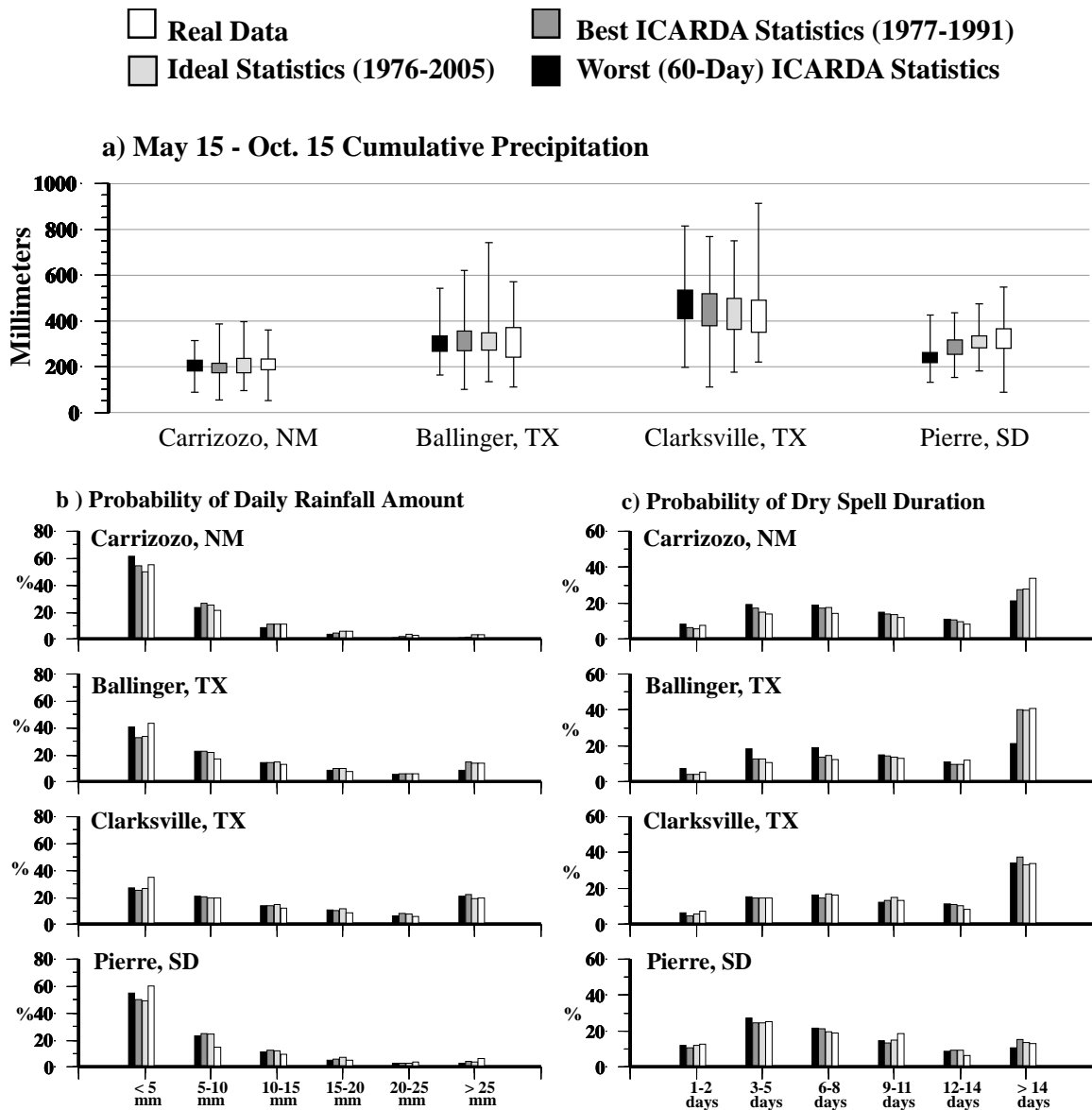


Fig. 6. a) Bar and whisker diagrams dividing real and generated May 15-Sept. 15 precipitation distributions into equal 33rd percentiles, for the 4 USHCN stations in Fig. 4. White bars indicate real distributions, gray shaded bars indicate distributions generated under the I,B, and W monthly statistics conditions. b) Probability distribution of real and generated daily rainfall amounts at each station during May 15-Sept. 15, using the same shading scheme in (a). c) Probability distribution of real and generated days between rain for each station during May 15-Sept. 15.

that the effects of the random errors in bi-weekly precipitation climatology seen in Figs. 5a-f may cancel in the calculation of longer-term precipitation statistics.

Figures 6b and c correspond to the green and gold bar graphs on the application's Precipitation Tab (Fig. 2u,v), calculated over the May 15-Oct. 15 summer growing season. Figure 6b was formed by sorting each of the four USHCN sites

daily rainfall totals into one of six bins, and shows the results of this sorting in both the real data and the artificial data generated under the 'I', 'B', and 'W' statistics conditions. Those figures show qualitative agreement between the distribution of generated and real daily rainfall amounts at all four sites, and each of the three statistics conditions. Figure 6c was formed by calculating the percentage of dry days that

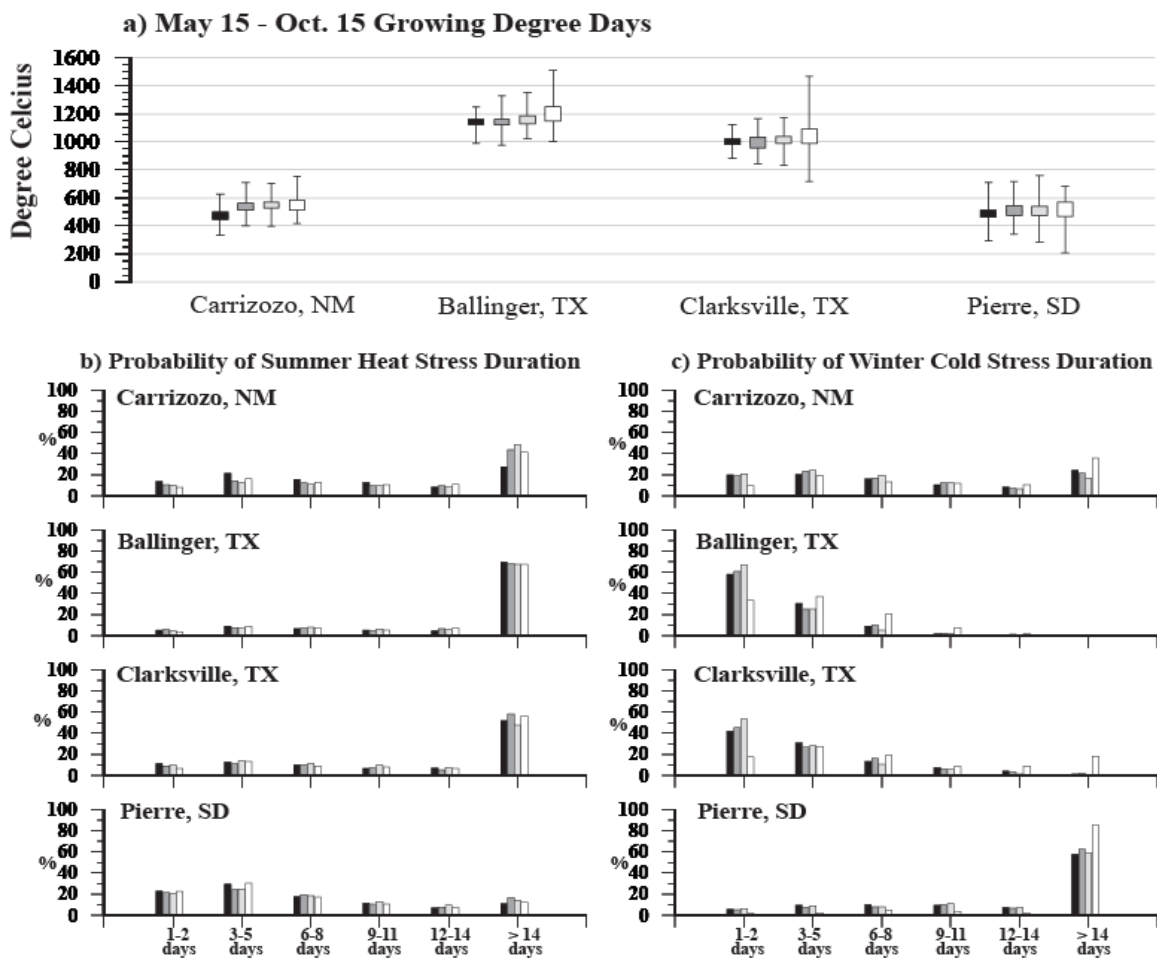


Fig. 7 a) As in Fig. 6a for May 15-Sept. 15 growing degree days calculated with an 18.33 °C temperature threshold. b) The probability that daily maximum temperature will exceed 30 °C in consecutive runs of 1-2, 3-5, 6-8, 9-11, 12-14, or more than 14 days in the real and generated data during May 15-Sept.15. c) The probability that daily minimum temperature will be below 0 °C in consecutive runs of 1-2, 3-5, 6-8, 9-11, 12-14, or more than 14 days in the real and generated data during Jan. 1-Mar. 15.

occurred in runs of 1-2, 3-5, 6-8, 9-11, 12-14, or more than 14 days in the real and generated data. Carrizozo shows a lower than observed frequency of 14+ day dry spells in the generated data under the I, B and W monthly statistics conditions. Ballinger's generated data shows a higher incidence of dry spells shorter than 9 days and a lower incidence of 14+ day dry spells. Clarksville and Pierre show closer agreement in dry run duration between the real data and the generated data under all three statistics conditions.

5.3 Seasonal Temperature Climatology

The Temperature Tab displays an exceedance curve for seasonal growing degree days (Fig. 2w), and bar charts showing the probability of heat (Fig. 2y) and cold stress

periods (Fig. 2z) of various duration. Figure 7a is Fig. 6a's counterpart for summer growing degree days (GDD) calculated with an 18.33 °C (65.0 °F) temperature threshold. As in Fig. 6a, the real data's site-to-site GDD variation is followed by the generated data under the I, B, and W monthly statistic conditions; e.g., the increases in generated GDD at Ballinger and Clarksville relative to Carrizozo and Pierre. But the central terciles for generated GDD at Ballinger and Clarksville are negatively biased relative to real growing degree days. Also, the height of the three generated distributions at those warmer sites are noticeably shorter than that of the real distributions, which shows that the weather generator produced less interannual variance in summer GDD than is found in the real data.

Figure 7b corresponds to the application's red bar graphs on the Temperature Tab, calculated over the May 15-Oct. 15 summer growing season. Those bar graphs are analogous to the Fig. 6b bar graphs, but show the percentage of hot days that occurred in the same six classes of run duration, in both the real and generated data. In the Fig. 7b test a day was considered hot when T_{\max} was greater than 30 °C (86 °F). In the application that heat stress threshold can be adjusted over a 15-45 °C range (Fig. 2h). The GEM6 output generated from Carrizozo monthly statistics under W conditions shows a lower probability of 14+ day hot spells than is found in the real data or the I and B generated data, and higher probabilities of shorter-term heat stress. At the other three stations the length of summer hot spells are in fairly close agreement in the generated and real data in each of the six duration categories, under each of the three generated data conditions.

Figure 7c is similar to Fig 7b, but shows the occurrence of cold runs that occurred in the real and generated data during January 1–March 15. Comparable cold run probabilities are displayed on the blue bar graphs on the application's Temperature Tab (Fig. 2i). The Fig. 7c test considered a day to be cold when T_{\min} was less than 0 °C (32 °F), but similar cold stress thresholds can be adjusted by the user (Fig. 2h) in the application. Carrizozo shows a higher probability of I, B, and W-generated short-term (< 9 day) cold spells, but lower probabilities of 14+ day dry spells than is found in the real data. At Ballinger and Clarksville all generated data shows a clearly higher incidence of 1-2 day cold spells, but lower incidences of longer term cold periods compared to the real data. At Pierre there is a consistently lower than observed probability of 14+ day cold spells in the I, B, and W-generated T_{\min} values.

The tendency for weather generators to produce lower than observed incidences of extreme daily temperature conditions, and Ballinger's and Clarksville's reduced I, B, and W summer GDD variance in Fig. 7a, might be traced to how stochastic streams of daily temperature are generated. Generators such as GEM6 randomly vary daily temperature about a station's mean annual temperature cycle, with day-to-day persistence determined by the A and B matrix values. Those values are in turn defined by correlations between daily T_{\min} , and T_{\max} at 0 and 1 day lags. But because those correlations are calculated over all of the available data, they reflect the general short-term persistence traits of

day-to-day weather variability, and not the persistence of abnormal periods. Thus, for example, weather generators are unlikely to produce the abnormally long runs of daily temperatures that produce high GDD values during summer growing seasons marked by drought. Distributions of generated GDD that lacked these extreme values would have less variance than their real counterparts.

5.4 Summary and Suggestions for the Application's Use

The Figs. 5-7 tests show that monthly statistics derived from smaller samples of daily data can lead to biases in the application's generated seasonal climate distributions. In Fig. 6a smaller data samples lead to increasing bias in generated summer precipitation at Clarksville and Pierre. In Fig. 7a Carrizozo's Worst case summer GDD is negatively biased relative to the observed distribution and the generated Ideal and Best-case distributions. Although Figure 6a shows distributions of seasonal precipitation that are reproduced reasonably well by the weather generator under each data condition, Figs. 5a-f shows bi-weekly precipitation climatology errors that increase under the Worst case conditions. While the effects of those random short-term errors might cancel in calculating precipitation statistics over 5 month periods, they may have a greater effect over sub-seasonal periods. As the application can calculate over periods as short as 10 days, shorter-term precipitation statistics might have random errors that increase as the daily data sample size decreases.

The method suggested here for identifying errors related to data sampling is based on the assumption that they may not be consistently evident at neighboring station locations. In operating the application the user selects a location by left-clicking that location on the large map in the application's upper-left corner (Fig. 2a). The nearby stations whose Fourier parameters will be used to estimate the location's GEM6 parameter set will then flash in sequence. *In practice, the user should always compare the application's results for a location with the corresponding results for each of those nearby stations. The user should also compare each of the nearby station's results with one another.* For example, in the application's GUI display the annual cycles for the probability of heat and cold stress might be compared (Fig. 2f), as might the probability of exceedance curves for precipitation (Fig. 2t) and growing degree days (Fig. 2w). If these comparisons show one station's results to differ clearly from the

remaining stations, that station's GEM6 parameters may have been derived from biased statistics. In that case the user should consider using results derived only from one of the remaining nearby stations, or selecting a nearby location that does not include the outlier station in the averaging process. But if the nearby station's results are similar with minor variation, the application may produce reasonable estimates of an area's climatology through the inverse-distance averaging algorithm. However, this averaging process cannot compensate for the inherent shortcomings of the weather generator, e.g., the insufficient interannual variance in GDD in some areas (Fig. 7a), which exist under even Ideal monthly statistics conditions.

6. Generation of secondary weather variables.

Secondary daily meteorological variables were derived here using parameterization relationships drawn from FAO-56 (Allen et al., 1998). For reference to the generation of those variables, see the Fig. 8a flow chart.

6.1 Dew Point Temperature

Daily dew point temperatures were estimated from daily minimum temperatures using the following parameterization scheme (FAO-56 Eq. 6-6):

- $T_{dew} = T_{min} - 2^{\circ}C$ for locations in arid areas,
- $T_{dew} = T_{min}$ elsewhere. (2)

Arid locations were defined here as stations with a mean annual temperature of 18 °C or greater, and a mean annual number of wet days of 55 or less.

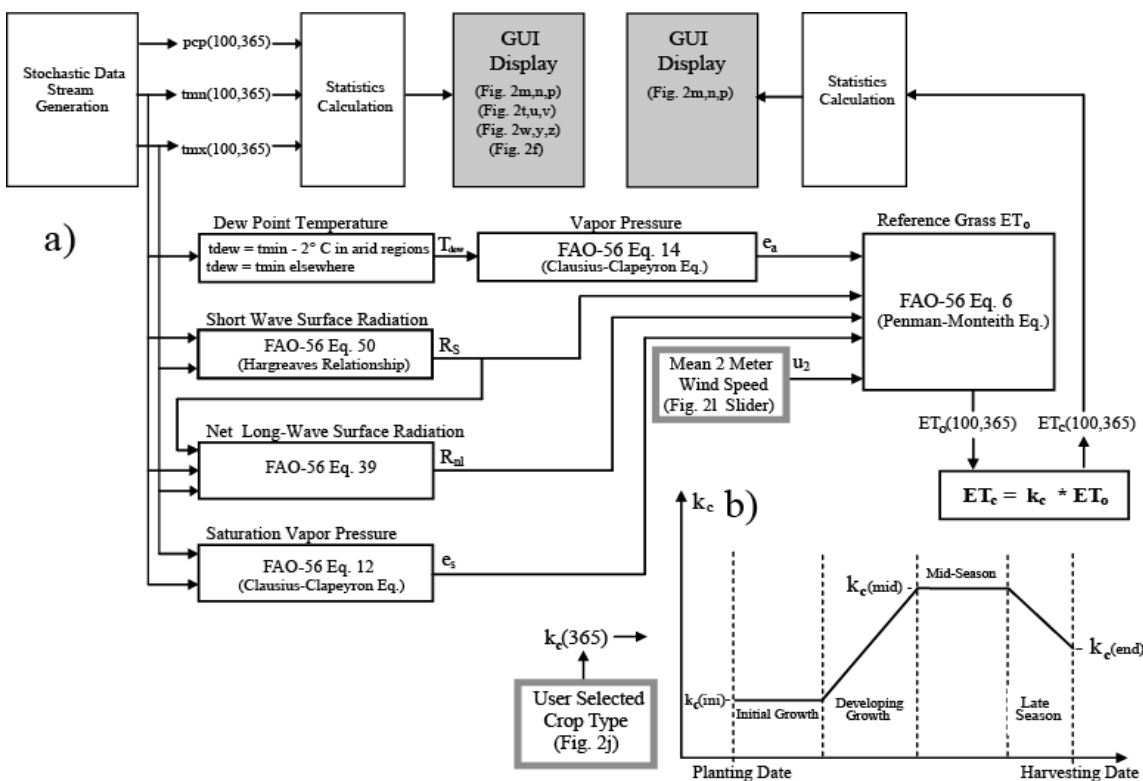


Fig. 8. a) Flow chart of the generation of daily secondary weather and crop evapotranspiration variables. Gray outlined elements indicate a user control, and gray shaded elements indicate a display marked in Fig. 2. b) Growing season coefficient profile for the FAO-56 single coefficient crop evapotranspiration method.

6.2 Shortwave radiation at the surface

Daily integrated shortwave surface radiation (R_s) was estimated using the Hargreaves radiation formula (FAO-56 Eq. 50):

$$R_s = k_{Rs} \sqrt{T_{max} - T_{min}} R_a \quad (3)$$

Where,

- k_{Rs} is an adjustment coefficient, assigned here as $0.175 \text{ } ^\circ\text{C}^{-0.5}$,
- T_{min} is daily minimum temperature ($^\circ\text{C}$),
- T_{max} is daily maximum temperature ($^\circ\text{C}$), and,
- R_a is the daily integrated shortwave radiation at the top of the atmosphere in units of Joules * $10^6 / (\text{met.}^2 * \text{day})$ (FAO-56 Eq. 21).

6.3 Vapor Pressure and Saturation Vapor Pressure.

Given daily minimum, maximum, and dew point temperatures, the vapor pressure and saturation vapor pressure are solved for using the Clausius-Clapeyron equation ($e^o(T)$: FAO-56 Eq. 11).

$$\text{Actual vapor pressure} = e_a = e^o(T_{dew}), \quad (4)$$

$$\text{Saturation vapor pressure} = e_s \\ e_s = 0.5 * (e^o(T_{max}) + e^o(T_{min})) \quad (5)$$

6.4 Net upwelling outgoing long-wave radiation (OLR) at the surface.

Net upwelling surface OLR was estimated using FAO-56 Eq. 39:

$$R_{nl} = \sigma \left[\frac{T_{min} + T_{max}}{2} \right] \left(0.34 - 0.14 \sqrt{e_a} \right) \left(1.35 \frac{R_s}{R_{so}} - 0.35 \right) \quad (6)$$

Where,

- σ is the Stefan-Boltzmann constant,
- T_{min} is daily minimum temperature in Kelvin,
- T_{max} is daily maximum temperature in Kelvin,
- R_s is the estimated daily shortwave radiation at the surface (Eq. 6)
- R_{so} is the estimated clear sky daily shortwave radiation at the surface (FAO-56 Eq. 37), and,
- e_a is the actual vapor pressure (Eq. 7).

6.4 Reference grass evapotranspiration

The FAO-56 method for deriving evapotranspiration rates for various crops is based on the estimation of reference evapotranspiration rates over a hypothetical grass surface (FAO-56 Chapter 4). Daily reference grass ET rates are calculated using the FAO-56 Penman-Monteith equation (FAO-56 Eq. 6).

$$ET_o = \frac{0.408 \Delta \left((1 - \alpha) R_s - R_{nl} - G \right) + \gamma \frac{900}{T} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (7)$$

Where,

- Δ is the slope of the saturation vapor pressure at the mean daily temperature (FAO-56 Eq. 3-3),
- α is the albedo of the hypothetical grass surface (=0.23),
- R_s is the shortwave solar radiation at the Earth's surface (Eq. 3),
- R_{nl} is the net upwelling outgoing long-wave radiation (OLR) at the surface (Eq. 6),
- G is the soil heat flux density (FAO-56 Eq. 5-2, with an assumed Leaf Area Index of 2.8),
- γ is the Psychrometric constant (FAO-56 Eq. 8),
- T is the daily mean (i.e., $0.5(T_{max} + T_{min})$) temperature in Kelvin,
- u_2 is the mean wind speed at 2 meters, set in the application via a slider control on the Crop ET tab (Fig. 21),
- e_a is the actual vapor pressure (Eq. 4), and,
- e_s is the saturation vapor pressure (Eq. 5).

6.5 Crop evapotranspiration

The application derives crop evapotranspiration rates over arbitrarily defined periods for a number of crops listed in the selection box at the top of the 'Crop ET' Tab (Fig. 2j). These crop ET rates (ET_c) are derived from a location's derived reference grass ET rates (ET_o) using the FAO-56 single crop coefficient method (FAO-56 Eq. 58).

$$ET_c = k_c ET_o \quad (8)$$

Over the growing season crop ET is derived from the reference grass rates using k_c values drawn from a growing season coefficient profile (Fig. 8b). That coefficient profile is in turn derived from three k_c values defined during an initial crop growth period, a mid-season period, and an end of season value.

7. Downloading, Operating Requirements and Instructions.

The ICARDA Agro-Climate Tool can be downloaded at:

<http://www.lbk.ars.usda.gov/WEWC/icarda.aspx>.

The application should be installed on a Windows PC with a Pentium III or better microprocessor and at least 230 Mbytes of available hard disk space. Monitor screen resolution should be at least 1024 X 768 pixels but no more than 1920 X 1440 pixels. The bright

yellow text on the graphical user interface provides basic operating instructions. To access help for a specific control or graph, left mouse click on that object and hit 'F1'. More detailed instructions for the application's use can be found by left single-clicking on 'Instructions' on the application's upper left corner.

Acknowledgements

Thanks to Robert Lascano, Terry Howell, and Steve Evett for suggestions and preliminary review comments. All figures were produced using Generic Mapping Tools (Wessel and Smith, 1995).

References

- Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. *Crop evapotranspiration: Guidelines for computing crop water requirements*. FAO Irrig. and Drain. Paper No. 56, Food and Agriculture Organization of the United Nations, Rome, Italy. 300 p.
(<http://www.fao.org/docrep/X0490E/x0490e00.htm>)
- Box, G.P. and G.M. Jenkins. 1978. *Time series analysis: forecasting and control*. Holden-Day, Inc. Oakland, 818 p.
- De Pauw, E. 2000. Drought early warning systems in West Asia and North Africa . In: D.A. Wilhite, M.V.K. Sivakumar, and D. A Woods, Eds. Early Warning Systems for Drought Preparedness and Drought Management. Proceedings of an Expert Group Meeting held in Lisbon, Portugal, 5-7 September 2000. World Meteorological Organization, Geneva. pp 65-85.
- De Pauw, E. 2007. Principal biomes in Central Asia. In: R. Lal, M. Suleimenov, B.A. Stewart, D.O. Hansen, and P. Doraiswamy, Eds. Climatic Change and Terrestrial Carbon Sequestration in Central Asia. Taylor and Francis, London. pp 3-24.
- Easterling, D. R., T. R. Karl, J. H. Lawrimore, and S. A. Del Greco, 1999: United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data for 1871-1997. ORNL/CDIAC-118, [NDP-070](#). Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Hanson, C.L., A. Cummings, D.A. Woolhiser, and C.W. Richardson. 1994. *Microcomputer program for daily weather simulation in the contiguous United States*. U.S. Department of Agriculture, Agricultural Research Service, ARS-114.
- Mendenhall, W., D.D. Wackerly, and R.L. Scheaffer. 1990. *Mathematical Statistics with Applications*. PWS-Kent, Boston, 818 p.
- National Climatic Data Center. 1994. Global Daily Summary: temperature and precipitation, 1977-1991. CD-ROM Version 1.0, Asheville, North Carolina.

National Climatic Data Center. 2002. The Global Daily Climatology Network (GDCN). CD-ROM Version 1.0, Asheville, North Carolina.

(<http://www.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html>).

National Weather Service, cited 2000: Cooperative Observer Program (COOP).

National Weather Service, Silver Spring, MD. [Available online at www.nws.noaa.gov/om/coop/Publications/coop.PDF].

Ryan, J., E. De Pauw, H. Gomez, and R. Mrabet, 2006. Drylands of the Mediterranean Zone: Biophysical resources and cropping systems. *In: G.A. Peterson, P.W. Unger and W.A. Payne Eds. Dryland Agriculture*. 2nd Ed. American Society of Agronomy Monograph Series No. 23. American Society of Agronomy. Madison, WI.. pp 577-624.

USGS (United States Geological Survey), cited 2006: Earth Resource Observation and Science (EROS),, Sioux Falls, SD. [<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>].

Wessel, P. and W.H.F. Smith, 1995: New version of the Generic Mapping Tools Released. *EOS, Trans. Amer. Geophys. Union*, **76**, pg. 329.

Wilks, D.S. 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93: 153-169.