Environmental Modelling & Software 24 (2009) 655–664

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Using MODAWEC to generate daily weather data for the EPIC model

Junguo Liu [a,*], Jimmy R. Williams [b], Xiuying Wang [b], Hong Yang [a]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Duebendorf, Switzerland
[b] Texas Agricultural Experiment Station, Blackland Research and Extension Center, 720 E. Blackland Road, Temple, TX 76502, USA

ARTICLE INFO

Article history:
Received 7 April 2008
Received in revised form
18 October 2008
Accepted 19 October 2008
Available online 22 November 2008

Keywords:
Daily weather generator
Crop yield
Evapotranspiration

ABSTRACT

Although the EPIC model has been widely used in agricultural and environmental studies, applications of this model may be limited in the regions where daily weather data are not available. In this paper, a stand-alone MODAWEC model was developed to generate daily precipitation and maximum and minimum temperature from monthly precipitation, maximum and minimum temperature, and wet days. A case study shows that the crop yields and evapotranspiration (ET) simulated with the generated daily weather data compare very well with those simulated with the measured daily weather data with low normalized mean square errors (0.008–0.017 for crop yields and 0.003–0.004 for ET). The MODAWEC model can extend the application of the EPIC model to the regions where daily data are not available or not complete. In addition, the generated daily weather data can possibly be used by other environmental models. Associated with MODAWEC, the EPIC model can play a greater role in assessing the impacts of global climate change on future food production and water use.

Software availability

Name of software: MODAWEC
Program language: FORTRAN
Developer: Swiss Federal Institute of Aquatic Science and Technology and Texas Agricultural Experiment Station
Contact address: Junguo Liu, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Ueberlandstrasse 133, CH-8600, Duebendorf, Switzerland
Software required: Windows 2000, XP, or Vista
Availability: Available to researchers free of charge on request to the corresponding author.

1. Introduction

Over the past two decades, the Environmental Policy Integrated Climate (EPIC, originally known as Erosion Productivity Impact Calculator) model has played an important role in the agricultural and environmental studies in the U.S. and in the other regions of the world. The EPIC model is a field-scale model that is designed to simulate drainage areas characterized by homogeneous weather, soil, landscape, crop rotation, and management system parameters. It was first developed in 1981 to support assessments of soil erosion impacts on soil productivity in the U.S. (Williams et al., 1984). Since then, it has continuously been developed by integrating and improving a number of additional functions including water quality, atmospheric CO_2 change, and carbon cycling routines. The model has been applied in a wide range of studies in agriculture, meteorology, and environment, e.g. crop growth and yield (Williams et al., 1989; Easterling et al., 1996), impacts of climate change (Easterling et al., 1992; Brown and Rosenberg, 1997), nutrient cycling and nutrient loss (Jackson et al., 1994; Pierson et al., 2001), wind and water erosion (Potter et al., 1998; Bhuyan et al., 2002), pesticide losses (Sabbagh et al., 1991; Williams et al., 1992), impacts of irrigation on crop yields (Cabelguenne et al., 1995; Rinaldi, 2001), soil temperature (Potter and Williams, 1994; Roloff et al., 1998), soil carbon sequestration (Lee et al., 1996; Potter et al., 2004), and economic–environmental analysis (Bernardo, 1993a,b; Kurkalova et al., 2004). Partly due to its good performance, the EPIC model has been applied in several regional, national and even global assessments. For example, a "spatial EPIC" system (Priya and Shibasaki, 2001) was developed to assess the national crop productivity in India. GIS-based EPIC models, which integrate EPIC with a geographic information system (GIS), were also used to study crop yield with high spatial resolutions for China (Liu et al., 2007a), for Africa (Liu et al., in press) and for the entire world (Tan and Shibasaki, 2003; Liu et al., 2007b, 2008; Liu, 2009).

Daily weather data are needed for the simulation of most processes in the EPIC model, but such data are often not available or not complete in many parts of the world. For example, so far, one of the most comprehensive daily weather data products, the Global Surface Summary of the Day produced by the National Climatic Data Center (NCDC), covers historical data of over 10,000 stations

* Corresponding author. Tel.: +41 448235012; fax: +41 448235375.
E-mail addresses: water21water@yahoo.com, Junguo.liu@eawag.ch (J. Liu).

1364-8152/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2008.10.008

from 1929 to the present, with data from 1973 to the present being the most complete (http://www.ncdc.noaa.gov). For this dataset, there is an uneven distribution of meteorological stations among countries with sparse stations in many underdeveloped countries. Furthermore, the daily data are often not complete with many missing data in individual stations. In addition, NCDC does not provide projected future weather data for these stations.

When not available or not complete, the daily weather data can be generated with EPIC's built-in "weather generator" (WXGEN). WXGEN incorporates a first-order Markov chain technique for a wet or dry day decision. When a wet day is generated, a skewed normal distribution is used to generate the amount of daily precipitation. WXGEN first independently generates precipitation for a day. Maximum temperature, minimum temperature, solar radiation and relative humidity are generated on the presence or absence of rain for the day. Daily wind speed is generated independently. Detailed description of WXGEN can be found in Sharpley and Williams (1990). The inputs to WXGEN are several monthly statistics taken from long-term daily weather records. Monthly statistics such as monthly skew coefficient and monthly probability of wet day after dry day or wet day are difficult to obtain without daily weather data. When the necessary monthly statistics are available, WXGEN is very useful in simulating daily weather sequences that have statistical properties similar to those of measured weather in the same region. It can provide any number of equally likely weather sequences for use in evaluating management strategies under varying climatic conditions. Also, it can repeat the same weather sequence of any length (hundreds of years) as many times as needed in evaluating various management strategies under the same climatic conditions. However, since WXGEN is a stochastic model, the generated weather sequences do not resemble measured weather records year to year, although their long-term statistical properties are similar. Daily time step models like EPIC require daily weather data, but these data cannot be generated accurately for individual years by WXGEN.

Monthly weather data are easier to obtain than daily weather data. For example, monthly precipitation, maximum and minimum temperature, and wet days are available on a global scale with a spatial resolution of 30 arc-min (about 50×50 km in each grid cell near the equator) for 1901–2000 through the Climatic Research Unit (CRU) at the University of East Anglia (Mitchell et al., 2004). The Tyndall Centre for Climate Change Research (TYN) from the same university provides monthly variations of the above climate data for 16 different climate scenarios for 2001–2100 (Mitchell et al., 2004). These monthly data are valuable for conducting past, current and future global environmental assessments. However, they cannot be used directly by EPIC because the model operates on a daily time step. Since only monthly weather is available in many locations, there is a need for a method for converting monthly data to daily data.

The purpose of this study is to develop a stand-alone weather generator MODAWEC (*MO*nthly to *DA*ily *WE*ather *C*onverter) for the EPIC model (EPIC0509). As our main interests are crop yield and crop water use, one important objective of model development is to generate reliable daily weather data for the EPIC model to simulate crop yield and crop evapotranspiration (ET). A case study is provided to test the reliability.

2. The MODAWEC model

The MODAWEC model converts monthly precipitation (in mm) and maximum and minimum temperature (in °C) to daily values while preserving the monthly totals and averages. The main inputs of the MODAWEC model include monthly precipitation, monthly wet days, and monthly maximum and minimum temperature in each year. The outputs are daily precipitation, daily maximum temperature, and daily minimum temperature. The flowchart of the MODAWEC model is depicted in Fig. 1. According to the classification by Bannayan and Hoogenboom (2008), the MODAWEC model is a parametric weather generator because it uses precipitation as the driving variable. Precipitation occurrence and amount are generated independently and other variables (e.g. temperature) are then generated based on the stochastically generated precipitation. We do not apply a nonparametric approach such as the *k*-nearest neighbor (*k*-NN) approach (Bannayan and Hoogenboom, 2008). The *k*-NN approach needs not only monthly weather data but also observed historical daily data in a number of years as input. As discussed previously, the observed historical daily data are often not available.

2.1. Daily precipitation generation

To generate daily precipitation, a first-order Markov chain by Nicks (1974) is first used to define the day as wet or dry. Markov chain models are based on transitional probability matrices of various time steps. Most often, a first-order Markov chain implies preservation of statistical parameters and especially the first-order autocorrelation coefficient in the synthetic sequences (Sahin and Sen, 2001). Here, precipitation occurrence is assumed to follow a two-state (dry or wet), first-order Markov chain with two transition probabilities: the probability that a wet day follows a dry day and the probability that a wet day comes after a previous wet days. The probability of rain on a given day is conditioned on the wet or dry status of the previous day. The transition from one state (dry or wet) to the other (wet or dry) is governed by the two transition probabilities. In case of a wet day, a modified exponential distribution is used to give first approximations of the amount of daily precipitation. Final daily precipitation is obtained by correcting the initial estimates based on the given monthly precipitation. The process of daily precipitation generation is described below. All the equations in this section are obtained from Williams (1995).

On any given day, the input to the daily precipitation generator must include information as to whether the previous day is dry or wet. A random number (0–1) is generated and compared with the appropriate wet–dry probability. If the random number is less than or equal to the wet–dry probability, precipitation occurs on that day. Random numbers greater than the wet–dry probability give no precipitation. Since the wet–dry state of the first day is established (by assuming the first day is dry), the process can be repeated for the next day and so on throughout the simulation period. Since wet–dry probabilities are not available for monthly precipitation data, they can be estimated given the number of wet days. The probability of a wet day is calculated directly from the number of wet days:

$$P_{(\mathrm{w})} = \frac{n_{\mathrm{w}}}{n} \tag{1}$$

where $P_{(\mathrm{w})}$ is the probability of a wet day, n_{w} is the number of wet days, and n is the number of days in a month. The probability of a wet day after a dry day can be estimated as a fraction of $P_{(\mathrm{w})}$

$$P_{(\mathrm{w|d})} = b_1 \times P_{(\mathrm{w})} \tag{2}$$

where $P_{(\mathrm{w|d})}$ is the probability of a wet day following a dry day and b_1 is a fraction usually in the range of 0.6–0.9. The probability of a wet day following a wet day can be calculated directly by using the equation

$$P_{(\mathrm{w|w})} = 1.0 - b_1 + P_{(\mathrm{w|d})} \tag{3}$$

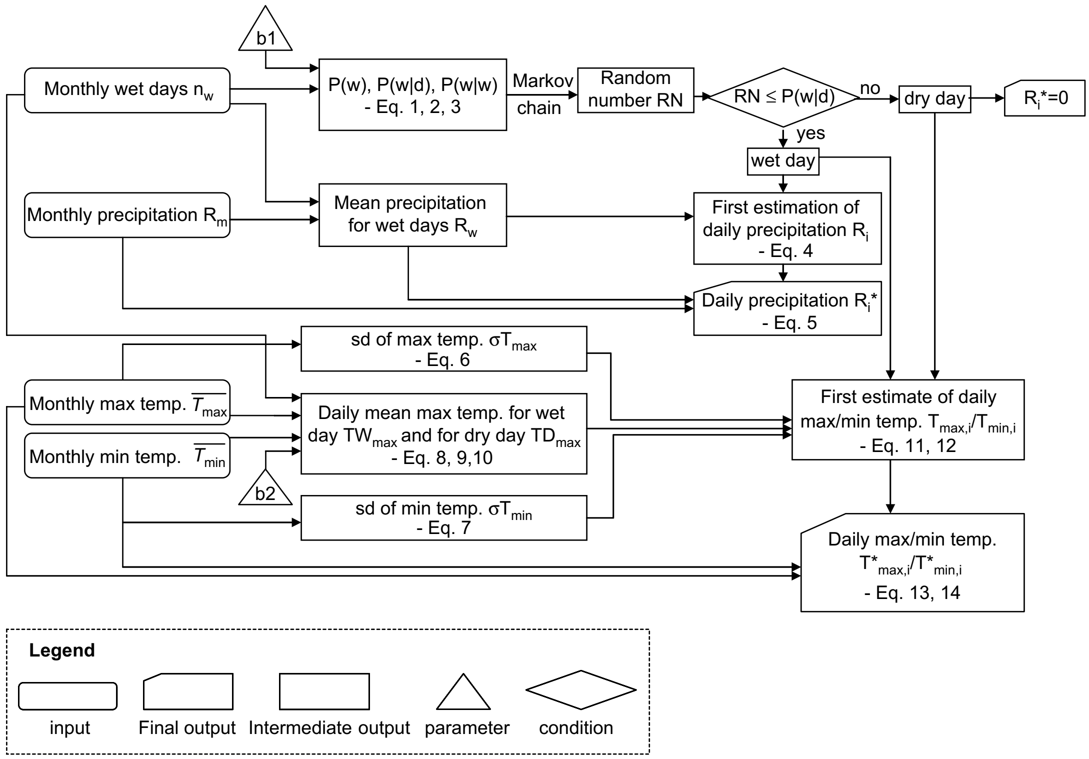

Fig. 1. Flowchart of the MODAWEC model.

where $P_{(w|w)}$ is the probability of a wet day after a wet day. When $b_1 = 1.0$, wet days do not affect probability of rainfall: $P_{(w|d)} = P_{(w|w)} = P_{(w)}$. Conversely, low b_1 values give strong wet day effects: $b_1 = 0.0$, $P_{(w|w)} = 1.0$, $P_{(w|d)} = 0$. Thus, b_1 controls the interval between rainfall events but has no effect on the number of wet days. For many locations, $b_1 = 0.75$ gives satisfactory estimates of $P_{(w|d)}$. In this paper, $b_1 = 0.75$ is used as default. Although Eqs. (2) and (3) may give slightly different probabilities from those estimated from rainfall records, they do guarantee correct simulation of the number of rainfall events.

In case of a wet day, daily precipitation is generated from a modified exponential equation:

$$R_i = R_w \times (-\ln(\mathrm{RN}))^{1.3} \tag{4}$$

where R_i is the daily precipitation on day i, R_w is the mean precipitation amount for wet days in a month, and RN is a uniform random number. The value of R_w is calculated by dividing the mean monthly precipitation (R_m) by the mean number of wet days (n_w) in each month. The R_i calculated in Eq. (4) are summed and corrected using the equation

$$R_i^* = R_i \times \frac{R_m}{S} \tag{5}$$

where R_i^* is the corrected daily precipitation on day i, R_m is the measured monthly precipitation, and S is the sum of the generated precipitation amounts for a month.

2.2. Daily maximum and minimum temperature generation

In MODAWEC, the model developed by Richardson (1981) was selected to give first approximations of daily temperature because it simulates temperature that is correlated with rainfall. The residuals of daily maximum and minimum air temperature are generated from a multivariate normal distribution. The multivariate generation model used implies that the residuals of maximum and minimum temperature are normally distributed and that the serial correlation of each variable may be described by a first-order linear autoregressive model. Final values of temperature are obtained by correcting the initial estimates using the average daily maximum and minimum temperature in a month.

The temperature model requires monthly means of maximum and minimum temperatures and their standard deviations as inputs. If the standard deviations are not available, the long-term observed extreme monthly minimums and maximums may be substituted. The model estimates standard deviation as 0.33 of the difference between the extreme and the mean for each month. If extreme temperatures are not available, MODAWEC estimates the standard deviations from the equations.

$$\sigma T_{max} = \max(0.5, 5.8 - 0.09 \times \overline{T}_{max}) \tag{6}$$

$$\sigma T_{min} = \max(0.5, 5.2 - 0.13 \times \overline{T}_{min}) \tag{7}$$

where for a month, σT_{max} is the standard deviation of daily maximum temperature, $\overline{T}_{max}$ is the average daily maximum temperature in the month, σT_{min} is the standard deviation of daily minimum temperature, and $\overline{T}_{min}$ is the average daily minimum temperature in the month.

Maximum temperature tends to be lower on rainy days. Thus, it is necessary to adjust the mean maximum temperature downward for simulating rainy day conditions. For the mean monthly maximum temperature ($\overline{T}_{max}$) this is accomplished by assuming that wet day values are less than dry day values by some fraction of the difference between $\overline{T}_{max}$ and $\overline{T}_{min}$:

$$\mathrm{TW}_{max} = \mathrm{TD}_{max} - b_2(\overline{T}_{max} - \overline{T}_{min}) \tag{8}$$

where TW_{max} is the daily mean maximum temperature for wet days, TD_{max} is the daily mean maximum temperature for dry days,

b_2 is a scaling factor ranging from 0.0 to 1.0, Choosing $b_2 = 1.0$ provides highest deviations on wet days and $b_2 = 0.0$ ignores the wet day effect. Observed data indicate that b_2 usually lies between 0.5 and 1.0. The default value of 0.5 is used in the MODAWEC model.

Since Eq. (8) gives lower mean maximum temperature values for wet days, a companion equation is necessary to slightly increase mean maximum temperature for dry days. The development is taken directly from the continuity equation

$$\overline{T}_{\max} \times n = \mathrm{TW}_{\max} \times n_{\mathrm{w}} + \mathrm{TD}_{\max} \times n_{\mathrm{d}} \tag{9}$$

where n is the number of days in a month, n_{w} is the number of wet days, and n_{d} is the number of dry days. The desired equation is obtained by substituting Eq. (8) into Eq. (9) and solving for $\mathrm{TD}_{\max}$.

$$\mathrm{TD}_{\max} = \overline{T}_{\max} + b_2(\overline{T}_{\max} - \overline{T}_{\min})\frac{n_{\mathrm{w}}}{n} \tag{10}$$

The use of the continuity equation guarantees that the long-term simulated value for mean maximum temperature agrees with the input value of $T_{\max}$. The first approximation of maximum and minimum temperature is estimated with the equations

$$T_{\max,i} = \mathrm{TX} + \sigma T_{\max} \times \mathrm{dT}_{\max} \tag{11}$$

$$T_{\min,i} = \mathrm{TN} + \sigma T_{\min} \times \mathrm{dT}_{\min} \tag{12}$$

where $T_{\max,i}$ and $T_{\min,i}$ are the first approximations of maximum and minimum temperature on day i, $\mathrm{dT}_{\max}$ and $\mathrm{dT}_{\min}$ are the standard normal deviates for maximum and minimum temperature, and $\mathrm{TX} = \mathrm{TW}_{\max}$ on wet days while $\mathrm{TX} = \mathrm{TD}_{\max}$ on dry days. Daily minimum temperature is assumed not to be affected by the wet/dry conditions. TN is equal to the mean monthly minimum temperature, or $\mathrm{TN} = \overline{T}_{\min}$.

The simulated $T_{\max,i}$ and $T_{\min,i}$ are corrected using the equations

$$T^*_{\max,i} = T_{\max,i} + \overline{T}_{\max} - \mathrm{SM}_1/n \tag{13}$$

$$T^*_{\min,i} = T_{\max,i} - n \times T_{\max,i} - T_{\min,i} \times \frac{\overline{T}_{\max} - \overline{T}_{\min}}{\mathrm{SM}_1 - \mathrm{SM}_2} \tag{14}$$

where $T^*_{\max,i}$ and $T^*_{\min,i}$ are the corrected maximum and minimum temperatures on day i, and SM_1 and SM_2 are the sums of $T_{\max,i}$ and $T_{\min,i}$ for a month.

2.3. Sensitivity analysis

While there are many methods for sensitivity analysis, variance-based sensitivity analysis is the most commonly used one. Variance-based sensitivity analysis is based on generated samples of input parameters from a defined probability distribution. With the generated samples, a model is evaluated to create output. The sensitivity of the output variance is analyzed in relation to the variation of the input parameters. The first-order sensitivity index (S_i) represents the sensitivity of output Y to a singular parameter X_i, and it is estimated as a ratio between the conditional variance of the expectation value $E(Y|X_i)$ and the unconditional variance of output Y (Schwieger, 2004) given by:

$$S_i = \frac{V(E(Y|X_i = x_i^*))}{V(Y)} \tag{15}$$

where $V(Y)$ is the total variance of output Y, X_i is an input parameter, $E(Y|X_i = x_i^*)$ is the expectation of Y conditional on X_i having fixed value x_i, $V(E(Y|X_i = x_i^*))$ is the conditional variance of estimated output Y where parameter X_i is fully fixed and others are varying. S_i represents the average output variance reduction that can be achieved when X_i becomes fully known and is fixed.

When dealing with non-addictive models, the estimation of higher order sensitivity indices is essential, but it is very computationally demanding. An efficient alternative is to compute the total sensitivity index (S_{Ti}). S_{Ti} represents the overall impact of parameter X_i on output Y (Schwieger, 2004) given by:

$$S_{Ti} = \frac{E(V(Y|X_{\sim i} = x^*_{\sim i}))}{V(Y)} \tag{16}$$

where $X_{\sim i}$ is all the parameters apart from X_i. $E(V(Y|X_{\sim i} = x^*_{\sim i}))$ is the conditional variance of Y when all parameters are fixed except X_i which is varying. S_{Ti} is the sum of all effect (first and higher order) involving the parameter X_i, and it is regarded as the expected fraction of the output variance that would remain unexplained if parameter X_i were unknown but all other parameters were known.

The extended Fourier amplitude sensitivity test (FAST) method is selected to calculate S_i and S_{Ti}. For the FAST method, Fourier decomposition is applied to obtain the fractional contribution of the individual parameters to the variance of the model output. The basis of the FAST approach is a transformation that converts a multidimensional integral over all the uncertain model input parameters to a one-dimensional integral, via a search curve that scans the whole parameter space (Saltelli and Bolado, 1998). The scanning is done so that each axis of the parameter space is explored with a different frequency. The classical FAST method initially developed by Cukier et al. (1970) can be used to estimate the effect of only one input parameter or the effect of all inputs varying together, but it is not efficient in addressing higher order interaction terms. The extended FAST method developed by Saltelli et al. (1999) can address higher order interactions between the input parameters.

We select the extended FAST method for sensitivity analysis mainly due to the many advantages of FAST over other sensitivity analysis methods. FAST represents one of the most elegant methods for sensitivity analysis. As a method for global sensitivity analysis, FAST is superior to the local sensitivity analysis methods mainly due to two reasons: first, it can apportion the output variance to the variance in the input parameters; and second, it can be used to fix the non-influential parameters at their midpoint or "nominal value" (Saltelli and Bolado, 1998). FAST is also superior to many other commonly used methods for global sensitivity analysis, e.g. standardized regression coefficients (SRC) (Draper and Smith, 1981) and standardized rank regression coefficients (SRRC) (Saltelli and Sobol, 1995). This is in part because that the FAST method is model independent, and it allows the determination of not only the individual effects of parameters, but also the cumulative interaction effect among parameters.

3. A test of the MODAWEC model

3.1. Site description and model parameterization

The weather, soil, and management data used in this test case were from a long-term experiment conducted at the Arlington Agricultural Research Station of the University of Wisconsin in the south central Wisconsin (43° 18′ N, 89° 21′ W). The station is located on an extended plain with 1%–2% slope on a Plano silt loam soil (fine-silty, mixed, mesic, Typic Argiudoll). The long-term experiment was established in 1958 in order to evaluate the response of continuous corn (*Zea mays* L.) to nitrogen treatments. Four treatments (T3, T5, T7 and T9) were selected in which nitrogen fertilizer was applied in the absence of liming. The treatments with lime application were not used here mainly due to the short liming period (Wang et al., 2005). All the treatments had the same tillage operations and planting/harvesting dates except for the nitrogen

Table 1
Nitrogen fertilizer application rates of four treatments.

Period	Nitrogen fertilizer application rate (kg N ha^{-1})			
	T3	T5	T7	T9
1958–1962	56	112	56	112
1963–1972	92	184	92	184
1973–1983	140	280	140	280
1984–1991	0	0	84	168

fertilizer application rates (Table 1). Corn was planted every year usually between the 1st and 4th weeks of May and harvested in the 4th week of October. Nitrogen fertilizer was generally applied 10 days prior to crop planting. A detailed description of the study site and treatment design can be found in Wang et al. (2005).

Long-term solar radiation, precipitation, maximum and minimum temperature, relative humidity and wind velocity have been measured on a daily basis for the period 1958–1991 (Wang et al., 2005). The soil profile was divided into five layers. Soil parameters of soil depth, bulk density, wilting point, field capacity, sand and silt content, soil pH and organic carbon content in each layer were based on measured values (see details in Wang et al., 2005). Parameters were set based on an automatic parameter optimization in Wang et al. (2005).

3.2. Simulation of crop yield and evapotranspiration (ET)

Daily potential increase in biomass is simulated using the Monteith's approach (Monteith, 1977). The daily biomass is adjusted for stress from water, temperature, nutrient and aeration (Williams et al., 1989). Crop yield is estimated by multiplying the above-ground biomass at maturity with a water stress adjusted harvest index. EPIC provides several methods to calculate potential ET (PET). The Penman–Monteith method (Monteith, 1965) is commonly used when all the weather variables of precipitation, maximum and minimum temperature, wind speed, relative humidity and solar radiation are available. When wind speed, relative humidity, and solar radiation data are not available, the Hargreaves method (Hargreaves and Samani, 1985) is an option, which estimates PET as a function of extraterrestrial radiation and temperature. The EPIC model computes transpiration from plants and evaporation from soil separately by an approach similar to that of Ritchie (1972). The potential evaporation and potential transpiration are first computed. Actual ET and transpiration may be limited by soil water deficits. Simulated crop yield is affected when different methods to estimate PET are used since the water stress constraint for crop growth is the ratio of actual transpiration to potential transpiration.

To test the MODAWEC model, crop yield and ET were simulated with four different sets of weather data (Run 1, 2, 3, and 4 in Table 2) for each of the four treatments in Table 1. In Run 1 and Run 2, daily precipitation, maximum temperature, minimum temperature, solar radiation, relative humidity and wind velocity from 1958 to 1991 were used. In Run 1, measured daily data on all these weather variables were used. In Run 2, daily precipitation, maximum temperature, and minimum temperature generated by MODAWEC were used with the measured daily data on other three variables. In Run 3 and Run 4, only daily data on precipitation, maximum temperature, and minimum temperature were used. For the three weather variables, measured daily data were used in Run 3, while daily data generated by MODAWEC were used in Run 4. Crop yield and ET calculated in Run 2 were compared with those calculated in Run 1. Crop yield and ET calculated in Run 4 were compared with those calculated in Run 3. The comparison will be shown in the next sections.

Table 2
Setup of four different runs in this study.

Run	Daily precipitation and maximum and minimum temperature	Daily solar radiation, relative humidity, and wind velocity	Method for PET	Results
Run 1	Measured data	Measured data	Penman–Monteith	Y_{P0}, ET_{P0}
Run 2	Generated using MODAWEC	Measured data	Penman–Monteith	Y_{P1}, ET_{P1}
Run 3	Measured data	Not used	Hargreaves	Y_{H0}, ET_{H0}
Run 4	Generated using MODAWEC	Not used	Hargreaves	Y_{H1}, ET_{H1}

3.3. Statistical tests for model performance

Several statistical indexes were used to test whether the crop yields and ET simulated with generated weather variables by MODAWEC are comparable with those simulated with measured weather variables. The indexes include coefficient of determination (r^2), slope and intercept of the regression function, normalized mean square error (NMSE) (Hanna, 1988), index of agreement (d) (Willmott, 1982). The normalized mean square error (NMSE) ranges from 0 to 1, and a value of 0 implies perfect agreement between two datasets. The index of agreement (d) ranges from 0 to 1, and a value of 1 implies perfect agreement.

3.3.1. Crop yield

Y_{P1} compares well with Y_{P0} for all the four treatments. All the full dots indicating the $Y_{P1} \sim Y_{P0}$ relation are scattered closely at the 1:1 lines (Fig. 2). For all the treatments, Y_{P1} and Y_{P0} agree well with each other with r^2 values between 0.83 and 0.87 (Table 3). All the slopes of the regression between Y_{P1} and Y_{P0} are not significantly different from 1.0 ($P < 0.0001$ for all treatments, F-test). The slopes are all close to 1.0. NMSE are between 0.008 and 0.013. The d values are all no less than 0.95. All the statistical indexes indicate excellent agreement between Y_{P1} and Y_{P0}.

Similarly, Y_{H1} compares well with Y_{H0} for all the four treatments. All the full dots indicating the $Y_{H1} \sim Y_{H0}$ relation are scattered closely at the 1:1 lines (Fig. 3). For all the treatments, Y_{H1} and Y_{H0} agree well with each other with r^2 values between 0.78 and 0.85 (Table 3). All the slopes of the regression between Y_{H1} and Y_{H0} are not significantly different from 1.0 ($P < 0.0001$ for all treatments, F-test). The slopes are between 0.94 and 0.99. NMSE are between 0.011 and 0.017. The d values are all no less than 0.94. All the statistical indexes indicate excellent agreement between Y_{H1} and Y_{H0}.

The above comparisons suggest that crop yields simulated with the daily precipitation and temperature data generated by MODAWEC compare very well with those simulated with the measured daily data on precipitation and temperature. To further compare between MODAWEC and WXGEN, we use the generated weather data by WXGEN as inputs to EPIC, and simulate crop yield with the Penman–Monteith and Hargreaves methods, respectively, for each of the four treatments. The comparison between these results and those from Y_{P0} and Y_{H0} shows poor agreement with r^2 ranged from 0.05 to 0.43 for Penman–Monteith method and r^2 ranged from 0.00 to 0.14 for Hargreaves method. The poor agreement is not surprising. WXGEN can generate any number of weather sequences that are equally likely to occur and that have the same statistical properties as the measured weather. However, none of these sequences will match the measured weather year by year. For example, WXGEN may generate dry conditions for a year that is very wet. Hence, WXGEN is a useful for long-term weather generation, but not for individual years. Instead, MODAWEC produces weather similar to the daily measured weather – similar enough that the simulated crop yields by EPIC are about the same as when

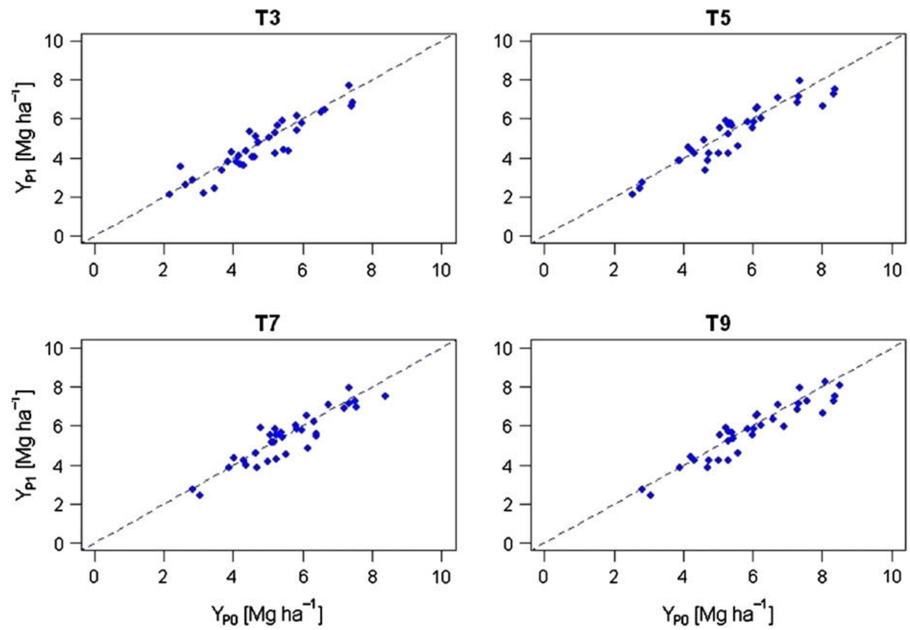

Fig. 2. The relation between Y_{P1} and Y_{P0}. The dashed line is the 1:1 line through the origin. Each point is for one specific year between 1958 and 1991.

measured daily weather is used. For the simulation of ET, similar conclusion can be drawn, and we will not repeat the comparison in later section.

Besides the above statistical tests, we also calculate the relative error, i.e. $(Y_{P1} - Y_{P0})/Y_{P0}$ and $(Y_{H1} - Y_{H0})/Y_{H0}$. The results show that about 85% of the relative errors for both the methods (Hargreaves and Penman–Monteith) in the four treatments are within ±20%, while over 95% of the relative errors are within ±30%. The high relative errors generally occur in the years when extreme weather occurs. One typical example is in the year of 1988. There are three subsequent days with maximum temperature over 38 °C in August. The high temperature results in damages to crop growth (e.g. crop yield is only 2.38 Mg/ha in T3). However, the extremely high temperatures are not completely captured by the MODAWEC model. The generated daily weather data only indicate one day with temperature higher than 38 °C in this month. As a result, the simulated yield is as high as 3.87 Mg/ha. It seems that, as most other traditional weather generators, the MODAWEC model has difficulties in accurately estimating extremes (temperature and precipitation included). Further research is needed to overcome this shortcoming, particularly when the model is used to support the simulation of extreme related processes such as flood or drought. According to Kilsby et al. (2007), the Neyman–Scott Rectangular Pulses (NSRP) model may be helpful to address this issue since it has been demonstrated to realistically reproduce extreme values for a number of sites in Italy and UK.

3.3.2. Evapotranspiration (ET)

ET_{P1} compares well with ET_{P0} for all the four treatments. All the full dots indicating the $ET_{P1} \sim ET_{P0}$ relation are scattered closely at the 1:1 lines (Fig. 4). For all the treatments, ET_{P1} and ET_{P0} agree well with each other with r^2 values between 0.83 and 0.84 (Table 3). All the slopes of the regression between ET_{P1} and ET_{P0} are not significantly different from 1.0 ($P < 0.0001$ for all treatments, *F*-test). The slopes range from 0.88 to 0.93. NMSE are all equal to 0.004. The *d* values are all equal to 0.96, close to 1 which indicates perfect agreement. All these statistical indexes show an excellent agreement between ET_{P1} and ET_{P0}.

Similarly, ET_{H1} compares well with ET_{H0} for all the four treatments. All the full dots indicating the $ET_{H1} \sim ET_{H0}$ relation are scattered closely at the 1:1 lines (Fig. 5). For all the treatments, ET_{P1} and ET_{P0} agree well with each other with r^2 values between 0.86 and 0.88 (Table 3). All the slopes of the regression between ET_{P1} and ET_{P0} are not significantly different from 1.0 ($P < 0.0001$ for all treatments, *F*-test). The slopes range from 0.90 to 0.91. NMSE are between 0.003 and 0.004. The *d* values are all equal to 0.97, close to 1 which indicates perfect agreement. All these statistical indexes show an excellent agreement between ET_{H1} and ET_{H0}.

Relative errors of ET are much smaller than those of crop yield. The results show that all the relative errors for both the methods (Hargreaves or Penman–Monteith) in the four treatments are within ±20%, while 99% of the relative errors are within ±15%. The smaller relative errors of ET are understandable. Crop yield can be highly affected by extreme weather. For example, high temperature can lead to reduction in yield. As a result, crop transpiration is lower. However, soil evaporation is generally a function of temperature, and it will increase with high temperature. Consequently, the sum of soil evaporation and crop transpiration, or ET, is less affected by extreme weather conditions.

Table 3
Statistical indexes indicating the extent of the agreement for $Y_{P1} \sim Y_{P0}$, $Y_{H1} \sim Y_{H0}$, $ET_{P1} \sim ET_{P0}$, and $ET_{H1} \sim ET_{H0}$.

Statistical index	$Y_{P1} \sim Y_{P0}$				$Y_{H1} \sim Y_{H0}$				$ET_{P1} \sim ET_{P0}$				$ET_{H1} \sim ET_{H0}$			
	T3	T5	T7	T9	T3	T5	T7	T9	T3	T5	T7	T9	T3	T5	T7	T9
r^2	0.85	0.86	0.83	0.87	0.81	0.85	0.78	0.84	0.83	0.84	0.84	0.84	0.86	0.88	0.88	0.88
Slope	0.94	0.93	0.94	0.93	0.97	0.98	0.94	0.99	0.88	0.92	0.93	0.93	0.90	0.90	0.91	0.91
Intercept	0.10	0.22	0.18	0.20	0.01	0.05	0.19	0.00	42	27	26	25	32	32	30	30
NMSE	0.013	0.011	0.010	0.008	0.017	0.012	0.014	0.011	0.004	0.004	0.004	0.004	0.004	0.003	0.003	0.003
d	0.96	0.96	0.95	0.97	0.95	0.96	0.94	0.96	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97

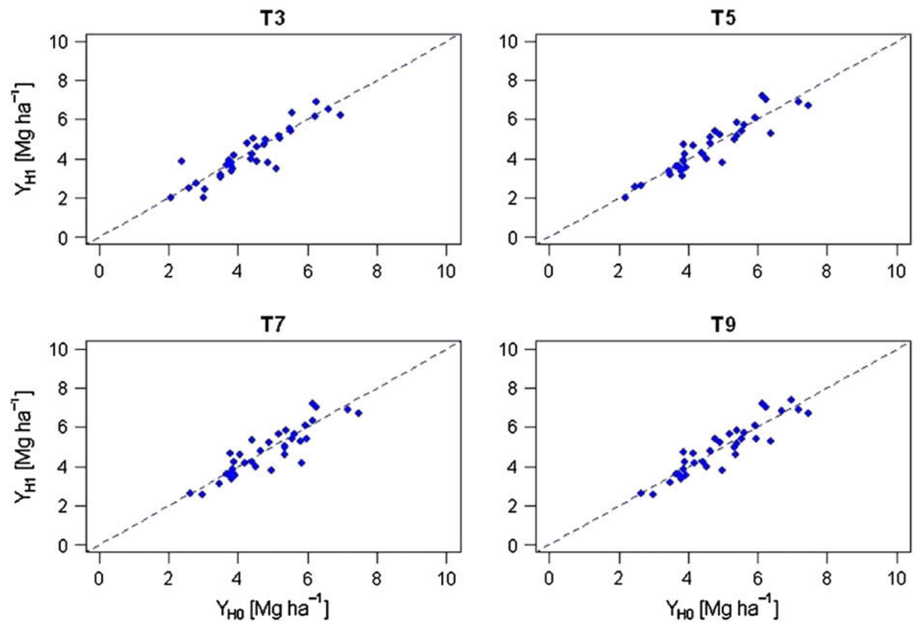

Fig. 3. The relation between Y_{H1} and Y_{H0}. The dashed line is the 1:1 line through the origin. Each point is for one specific year between 1958 and 1991.

The above comparisons suggest that ET simulated with the daily precipitation and temperature data generated by MODAWEC compares very well with that simulated with the measured daily data on precipitation and temperature, no matter which method (Penman–Monteith method or Hargreaves method) is used.

3.4. Sensitivity analysis

In this paper, sensitivity analysis is conducted with four steps:

1) Parameter sampling: the first step is to generate random parameter combinations of b_1 and b_2, the two parameters used in the MODAWEC model. A uniform distribution is assumed for both the parameters based on our experience. The ranges of parameters are 0.6–0.9 for b_1 and 0.5–1.0 for b_2. The sample size is 1000.
2) Execution of the MODAWEC model: the MODAWEC model is executed to calculate daily weather variables (i.e. maximum temperature, minimum temperature, and precipitation) for each combination of b_1 and b_2. With 1000 parameter combinations, we estimate 1000 sets of daily weather variables.
3) Execution of the EPIC model: crop yield and ET are calculated for each set of daily weather variables;
4) Calculation of sensitivity indices: the sensitivity indices are calculated based on input parameters (i.e. b_1 and b_2) and outputs (i.e. crop yield and ET) with the extended FAST method.

The parameter sampling and calculation of sensitivity indices are carried out with the "sensitivity" package for the R software. The sensitivity of simulated crop yield to the input parameters is demonstrated in Fig. 6 with the FAST first order and total order

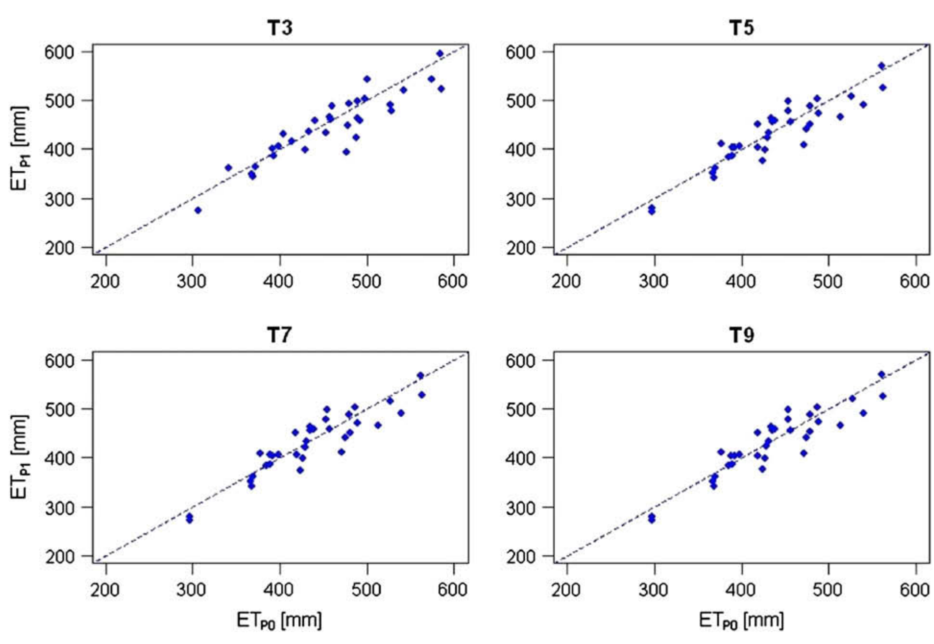

Fig. 4. The relation between ET_{P1} and ET_{P0}. The dashed line is the 1:1 line through the origin. Each point is for one specific year between 1958 and 1991.

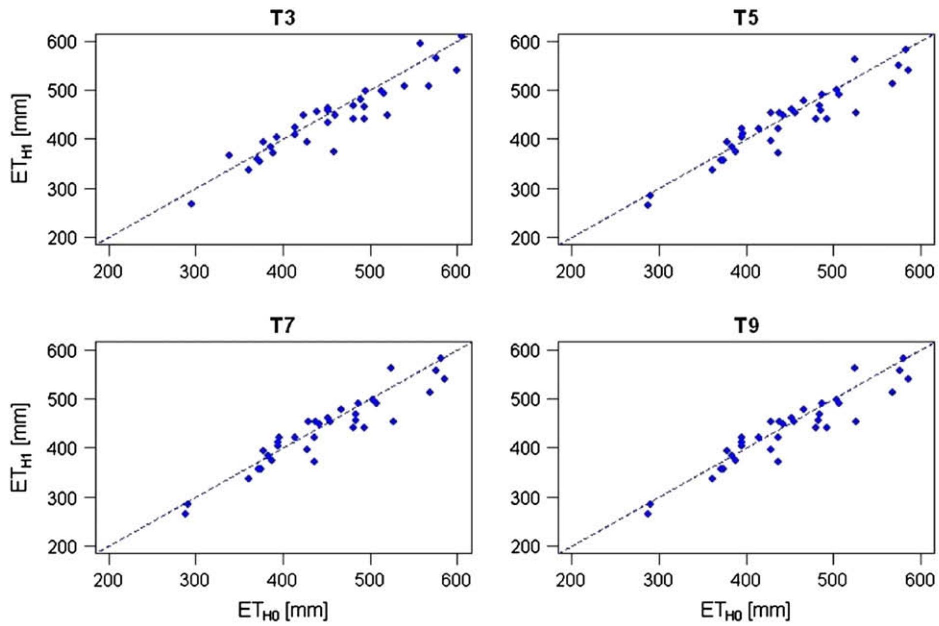

Fig. 5. The relation between ET_{H1} and ET_{H0}. The dashed line is the 1:1 line through the origin. Each point is for one specific year between 1959 and 1991.

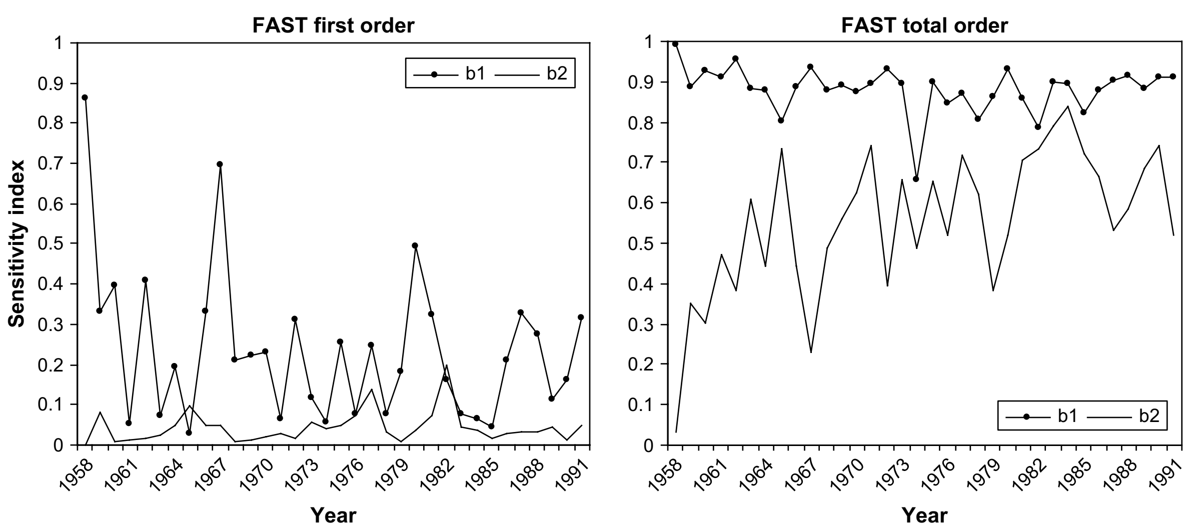

Fig. 6. Sensitivity indices for annual crop yield of corn for T7.

sensitivity indices. These sensitivity indices are calculated for T7 for Run 2 in Table 2 for illustration.

In general, the first order sensitivity index (S_i) for b_1 is much higher than that for b_2 (Fig. 6). S_i is 0.235 ± 0.185 for b_1 and 0.043 ± 0.039 for b_2 between 1958 and 1991. S_i for a particular parameter indicates the amount of variance that would be removed from the total output variance if the true value of that parameter were known. Hence, the comparison indicates that crop yield is more sensitive to parameter b_1 than b_2. This is so because b_1 affects storm interval and rainfall amount and b_2 affects temperature. Storm interval and rainfall amount largely determine runoff and infiltration (soil water available for plant use). Generally crop yield is much more sensitive to soil water than air temperature.

The sum of S_i for b_1 and b_2 is 0.279 ± 0.184 during 1958–1991. This means that, without considering the interaction between b_1 and b_2, the two parameters together can explain only about 30% of the output variance. Another obvious trend is that the FAST total order sensitivity index (S_{Ti}) is much higher than S_i (Fig. 6). This implies high interaction between b_1 and b_2. When the parameter interaction is taken into account, on average, b_1 can explain 88.1% of the output variance, while b_2 can explain 55.6% during 1958–1991. The high interaction is largely because that temperature is influenced by precipitation. Daily maximum temperature is calculated based on wet/dry conditions, as shown in Eqs. (8)–(11).

Sensitivity analysis is also conducted for crop yield and ET for all the treatments specified in Table 1. The results show the similar trends as described in the above illustration. In general, outputs are more sensitive to parameter b_1 than b_2. S_{Ti} is much higher than S_i, indicating high interaction between b_1 and b_2. Due to the similar conclusions drawn from each treatment, we only present sensitivity analysis for crop yield for T7 here in this paper.

4. Conclusions

A MODAWEC model has been developed to generate daily precipitation and maximum and minimum temperature with monthly weather data. The results show that the simulated crop yield or ET does not differ significantly when the measured daily weather data or the generated daily weather data by the MODAWEC model are used. The MODAWEC model enables the application of the EPIC model in regions where only monthly data are available. Particularly, for global level studies, daily weather data are often lacking in many regions, but complete historical monthly weather data on precipitation, maximum and minimum temperature and wetting days are available from the Climatic Research Unit (CRU) in the University of East Anglia for all grid cells with high spatial resolution. Furthermore, the MODAWEC model can contribute to assessing the impacts of global climate change on food production and water use. Future monthly weather data can be obtained from public available sources such as the Tyndall Centre for Climate Change Research in the University of East Anglia. A combined application of these monthly data with the EPIC and MODAWEC models can help simulate the impact of climate change on food production. Although the MODAWEC model is developed for the EPIC model, the generated daily weather data can also be used by other environmental models. The MODAWEC model and its source code are available free of charge by contacting the corresponding author of this paper.

The MODAWEC model has only been tested with four treatments at the Arlington Agricultural Research Station in the USA. Clearly further tests of the model in other locations are needed prior to wider applications. It also needs to be pointed out that the generated daily weather data may have poor correlation with the actual daily values, particularly for precipitation. The simulated wet days according to the Markov rainfall model may not be the actual wet days. This can lead to large differences between the generated and actual daily precipitation. One would not expect agreement on a daily basis since the rainfall occurrence and amounts are stochastic. However, agreement is guaranteed between the mean monthly simulated and measured values and the number of storms in each month. One would also expect close agreement between the standard deviations of simulated and measured daily rainfall amounts and maximum and minimum air temperatures. Although the simulated weather sequence is different from the measured sequence it should be just as valuable for most simulation projections. After all, the measured weather sequence will never be repeated in the future.

The MODAWEC model is developed to provide daily weather data in order to support the simulation of biogeochemical processes, e.g. crop yield and ET in this paper. We have demonstrated that the quality of the generated daily weather data is high enough for the simulation of crop yield and ET at the Arlington Agricultural Research Station. Nevertheless, there is a need to test whether the simulation results with the generated daily weather data are reliable for the simulation of other processes in the EPIC model, such as nutrient cycle, soil carbon sequestration, and wind and water erosion, etc.

Acknowledgements

This study was supported by the European Commission within the GEO-BENE project framework (Global Earth Observation–Benefit Estimation: Now, Next and Emerging, Proposal No. 037063). Special thanks are given to Monireh Faramarzi and Jafet Andersson from Eawag for their advice on sensitivity analysis in this paper.

References

Bannayan, M., Hoogenboom, G., 2008. Weather analogue: a tool for real-time prediction of daily weather data realizations based on a modified *k*-nearest neighbor approach. Environmental Modelling and Software 23 (6), 703.

Bernardo, D.J., 1993a. Economic and environmental impacts of water quality protection policies 1. Framework for regional analysis. Water Resources Research 29 (9), 3069.

Bernardo, D.J., 1993b. Economic and environmental impacts of water quality protection policies 2. Application to the central high plains. Water Resources Research 29 (9), 3081.

Bhuyan, S.J., Kalita, P.K., Janssen, K.A., Barnes, P.L., 2002. Soil loss predictions with three erosion simulation models. Environmental Modelling and Software 17 (2), 137.

Brown, R.A., Rosenberg, N.J., 1997. Sensitivity of crop yield and water use to change in a range of climatic factors and CO_2 concentrations: a simulation study applying EPIC to the central United States. Agricultural and Forest Meteorology 83 (3–4), 171.

Cabelguenne, M., Jones, C.A., Williams, J.R., 1995. Strategies for limited irrigations of maize in southwestern France – a modeling approach. Transactions of the American Society of Agricultural Engineers 38 (2), 507.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1970. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. The Journal of Chemical Physics, 3873.

Draper, N.R., Smith, H., 1981. Applied Regression Analysis. Wiley, New York.

Easterling, W.E., Chen, X., Hays, C., Brandie, J.R., Zhang, H., 1996. Improving the validation of model-simulated crop yield response to climate change: an application to the EPIC model. Climate Research 6 (3), 263.

Easterling, W.E., McKenney, M.S., Rosenberg, N.J., Lemon, K.M., 1992. Simulations of crop response to climate change: effects with present technology and no adjustments (the "dumb farmer" scenario). Agricultural and Forest Meteorology 59 (1–2), 53.

Hanna, S.R., 1988. Air quality evaluation and uncertainty. Journal of the Air Pollution Control Association 38, 406.

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1, 96–99.

Jackson, L.E., Stivers, L.J., Warden, B.T., Tanji, K.K., 1994. Crop nitrogen utilization and soil nitrate loss in a lettuce field. Fertilizer Research 37 (2), 93.

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., James, P., Smith, A., Wilby, R.L., 2007. A daily weather generator for use in climate change studies. Environmental Modelling and Software 22 (12), 1705.

Kurkalova, L., Kling, C.L., Zhao, J., 2004. Multiple benefits of carbon-friendly agricultural practices: empirical assessment of conservation tillage. Environmental Management 33 (4), 519.

Lee, J.J., Phillips, D.L., Dodson, R.F., 1996. Sensitivity of the US corn belt to climate change and elevated CO_2: II. Soil erosion and organic carbon. Agricultural Systems 52 (4), 503.

Liu, J., 2009. A GIS-based tool for modelling large-scale crop–water relations. Environmental Modeling & Software 24 (3), 411–422.

Liu, J., Fritz, S., van Wesenbeeck, C.F.A., Fuchs, M., You, L., Obersteiner, M., Yang, H. A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change. Global and Planetary Change, in press. doi:10.1016/j.gloplacha.2008.09.007.

Liu, J., Wiberg, D., Zehnder, A.J.B., Yang, H., 2007a. Modelling the role of irrigation in winter wheat yield, crop water productivity, and production in China. Irrigation Science 26 (1), 21–33.

Liu, J., Williams, J.R., Zehnder, A.J.B., Yang, H., 2007b. GEPIC – modelling wheat yield and crop water productivity with high resolution on a global scale. Agricultural Systems 94 (2), 478–493.

Liu, J., Zehnder, A.J.B., Yang, H., 2008. Drops for crops: modelling crop water productivity on a global scale. Global NEST Journal 10 (3), 295–300.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M., New, M., 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Tyndall Centre, Working Paper 55, Norwich, UK.

Monteith, J.L., 1965. Evaporation and Environment. In: Nineteenth Symposia of the Society for Experimental Biology. University Press, Cambridge, pp. 205–234.

Monteith, J.L., 1977. Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society B 281, 277–294.

Nicks, A.D., 1974. Stochastic generation of the occurrence, pattern, and location of maximum amount of daily rainfall. In: Proceedings Symposium on Statistical Hydrology. United States Department of Agriculture, Misc. Publication No. 1275: Tucson, AZ. 31 August–2 September 1971.

Pierson, S.T., Cabrera, M.L., Evanylo, G.K., Schroeder, P.D., Radcliffe, D.E., Kuykendall, H.A., Benson, V.W., Williams, J.R., Hoveland, C.S., McCann, M.A., 2001. Phosphorus losses from grasslands fertilized with broiler litter: EPIC simulations. Journal of Environmental Quality 30 (5), 1790.

Potter, K.N., Williams, J.R., 1994. Predicting daily mean soil temperatures in the EPIC simulation model. Agronomy Journal 86 (6), 1006.

Potter, K.N., Williams, J.R., Larney, F.J., Bullock, M.S., 1998. Evaluation of EPIC's wind erosion submodel using data from southern Alberta. Canadian Journal of Soil Science 78 (3), 485.

Potter, S.R., Atwood, J.D., Kellogg, R.L., Williams, J.R., 2004. An approach for estimating soil carbon using the national nutrient loss database. Environmental Management 33 (4), 496.

Priya, S., Shibasaki, R., 2001. National spatial crop yield simulation using GIS-based crop production model. Ecological Modelling 136 (2–3), 113.

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resources Research 17 (1), 182.

Rinaldi, M., 2001. Application of EPIC model for irrigation scheduling of sunflower in Southern Italy. Agricultural Water Management 49 (3), 185.

Ritchie, J.T., 1972. A model for predicting evaporation from a row crop with incomplete cover. Water Resources Research 8, 1204–1213.

Roloff, G., De Jong, R., Nolin, M.C., 1998. Crop yield, soil temperature and sensitivity of EPIC under central-eastern Canadian conditions. Canadian Journal of Soil Science 78 (3), 431.
Sabbagh, G.J., Geleta, S., Elliott, R.L., Williams, J.R., Griggs, R.H., 1991. Modification of epic to simulate pesticide activities. EPIC-PST. Transactions of the American Society of Agricultural Engineers 34 (4), 1683.
Sahin, A.D., Sen, Z., 2001. First-order Markov chain approach to wind speed modelling. Journal of Wind Engineering and Industrial Aerodynamics 89 (3–4), 263.
Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics and Data Analysis 26 (4), 445.
Saltelli, A., Sobol, I.M., 1995. About the use of rank transformation in sensitivity analysis of model output. Reliability Engineering and System Safety 50 (3), 225.
Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41 (1), 39–56.
Schwieger, V., 2004. Variance-based sensitivity analysis for model evaluation in engineering surveys. In: INGEO 2004 and FIG Regional Central and Eastern European Conference on Engineering Surveying, Bratislava, Slovakia. November 11–13, 2004.
Sharpley, A.N., Williams, J.R., 1990. Erosion/productivity impact calculator, I. Model documentation. USDA-ARS Technical Bulletin 1768.
Tan, G., Shibasaki, R., 2003. Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. Ecological Modelling 168 (3), 357.
Wang, X., He, X., Williams, J.R., Izaurralde, R.C., Atwood, J.D., 2005. Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC. Transactions of the ASAE 48 (3), 1041–1054.
Williams, J.R., 1995. The EPIC model. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado.
Williams, J.R., Jones, C.A., Dyke, P.T., 1984. A modelling approach to determining the relationship between erosion and soil productivity. Transactions – American Society of Agricultural Engineers 27 (1), 129.
Williams, J.R., Jones, C.A., Kiniry, J.R., Spanel, D.A., 1989. The EPIC crop growth model. Transactions of the ASAE 32, 497–511.
Williams, J.R., Richardson, C.W., Griggs, R.H., 1992. The weather factor: incorporating weather variance into computer simulation. Weed Technology 6, 731–735.
Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society 63 (11), 1309–1313.