Agricultural and Forest Meteorology 209–210 (2015) 49–58

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Creating long-term weather data from thin air for crop simulation modeling

Justin Van Wart[a], Patricio Grassini[a,*], Haishun Yang[a], Lieven Claessens[b,c], Andrew Jarvis[d], Kenneth G. Cassman[a]

[a] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583-0915, USA
[b] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), P.O. Box 39063, 00623 Nairobi, Kenya
[c] Soil Geography and Landscape Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[d] International Center for Tropical Agriculture (CIAT), Apartado Aéreo 6713, Cali, Colombia

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 18 February 2015
Accepted 28 February 2015
Available online 19 May 2015

Keywords:
Yield potential
Yeild variability
Weather data
Crop simulation model

ABSTRACT

Simulating crop yield and yield variability requires long-term, high-quality daily weather data, including solar radiation, maximum (T_{max}) and minimum temperature (T_{min}), and precipitation. In many regions, however, daily weather data of sufficient quality and duration are not available. To overcome this limitation, we evaluated a new method to create long-term weather series based on a few years of observed daily temperature data (hereafter called propagated data). The propagated data are comprised of uncorrected gridded solar radiation from the Prediction of Worldwide Energy Resource dataset from the National Aeronautics and Space Administration (NASA–POWER), rainfall from the Tropical Rainfall Measuring Mission (TRMM) dataset, and location-specific calibration of NASA–POWER T_{max} and T_{min} using a limited amount of observed daily temperature data. The distributions of simulated yields of maize, rice, or wheat with propagated data were compared with simulated yields using observed weather data at 18 sites in North and South America, Europe, Africa, and Asia. Other sources of weather data typically used in crop modeling for locations without long-term observed weather data were also included in the comparison: (i) uncorrected NASA–POWER weather data and (ii) generated weather data using the MarkSim weather generator. Results indicated good agreement between yields simulated with propagated weather data and yields simulated using observed weather data. For example, the distribution of simulated yields using propagated data was within 10% of the simulated yields using observed data at 78% of locations and degree of yield stability (quantified by coefficient of variation) was very similar at 89% of locations. In contrast, simulated yields based entirely on uncorrected NASA–POWER data or generated weather data using MarkSim were within 10% of yields simulated using observed data in only 44 and 33% of cases, respectively, and the bias was not consistent across locations and crops. We conclude that, for most locations, 3 years of observed daily T_{max} and T_{min} data would allow creation of a robust weather data set for simulation of long-term mean yield and yield stability of major cereal crops.

1. Introduction

Due to year-to-year fluctuation in weather patterns, long-term daily weather data, including solar radiation, temperature (maximum [T_{max}] and minimum [T_{min}]), and precipitation, are required to estimate crop yield potential and its variability using crop simulation models (Whisler et al., 1986; Boote et al., 1996; van Bussel et al., 2011 Van Wart et al., 2013a). Such estimates of yield potential and its variability are essential for analysis of food security, assessing impact of climate change on crop production, development and use of crop management decision-support tools, and to support and target agronomic research and policy. Depending on the degree of weather variability among years, at least 10–20 years of daily weather data are necessary for reliable estimates of mean yield potential and its inter-annual variability (van Ittersum et al., 2013 ; Van Wart et al. 2013a; Grassini et al., 2015). In many parts of the world, however, most weather stations only have a few years of daily weather records available and often not all of

Abbreviations: GridWD, gridded weather data; GenWD, generated weather data; RH, relative humidity; T_{min}, minimum temperature; T_{max}, maximum temperature; T_{dew}, dew point temperature; ETo, grass-based reference evapotranspiration; OWD, observed weather data; PWD, propagated weather data.

* Corresponding author. Tel.: +1 4024725554.
E-mail address: pgrassini2@unl.edu (P. Grassini).

http://dx.doi.org/10.1016/j.agrformet.2015.02.020

the variables necessary for crop model simulations are measured (*e.g.*, incident solar radiation). Unfortunately, many regions with limited availability of weather data (*e.g.*, Sub-Saharan Africa) are of greatest concern with regard to food security and vulnerability to climate change (Lobell et al., 2008). Hence, it is important to develop methods for generating reliable, long-term weather data for these regions where availability of weather data severely limits ability to perform robust assessments of yield gaps, and food security scenarios.

Gridded weather data (GridWD) or generated weather data (GenWD) have been used as alternatives in regions where observed weather data (OWD) are not available (Table 1). Crop simulation studies relying on GridWD and GenWD, however, have rarely compared simulated yields against simulations using OWD from weather stations located within the area of study. However, this is crucial because, in generating long-term weather data with global spatial coverage, sources of error can be incorporated into both GridWD and GenWD that can result in biased estimates of crop yield and its variability over time.

GridWD are typically derived by interpolation of observed weather data over space, or may also be derived from global climate models, to estimate daily or monthly weather data for each individual grid cell of land area (Kanamitsu et al., 2002; New et al., 2002). The quality of the estimation for a given grid cell depends on the density and distribution of the weather stations used in its derivation. Because both density and distribution are far from satisfactory in many regions of the world, derived GridWd in these regions are subject to a large degree of uncertainty. In fact, even in regions with an adequate density of weather stations, poor agreement has been found between simulated crop yields using GridWD *versus* simulations using OWD from a location within the same grid cell (Mearns et al., 2001; Baron et al., 2005). Regardless of whether the GridWD are derived through interpolation or from climate models, the bias in simulated yields using GridWD, relative to yields simulated using OWD, has been found to be unpredictable and inconsistent, having different sign and magnitude across locations for temperature and rainfall (Van Wart et al., 2013b).

A stochastic weather generator produces synthetic time series of daily weather data (GenWD) for as many years as specified for a location based on the statistical characteristics of historical daily or monthly OWD at that location (Hutchinson, 1987; Jones and Thornton, 2000; Hansen and Mavromatis, 2001; Mavromatis and Hansen, 2001). Models for generating stochastic weather data are typically developed in two steps: the first step is to model daily precipitation and the second step is to model or estimate the remaining variables of interest, such as daily T_{max} and T_{min}, solar radiation, humidity and wind speed. Even when decades of daily OWD are used to calibrate weather generators, they may perform poorly when compared to simulated crop yields based on OWD and typically underestimate inter-annual variation in crop model simulations (Semenov and Porter, 1995; Hammer et al., 2002). Likewise, though monthly means and variances of GenWD and OWD may be similar, short periods of extreme events, which are of particular importance for crop growth, yield and even crop failure, are typically not well represented in generated data (Kyselý and Dubrovsky, 2005; Semenov, 2008). While there are continuing efforts to improve weather generators, such efforts are constrained by the number of years and sites required for their parameterization (Baigorria and Jones, 2010; Rosenzweig et al., 2013).

Daily OWD with sufficient number of years to simulate long-term average crop yield and its variability (>10 years) are not available for many regions of the world. In contrast, short-term OWD of several years duration (typically <5 years) with daily maximum and minimum temperature (T_{max} and T_{min}, respectively) is often available for most regions. For example, in Africa there are a total of 1048 meteorological stations reporting at least 3 years of publically available weather data, but less than 12% of these stations have at least 15 years of OWD of adequate quality (missing <10% of total data and with no more than 30 data days missing consecutively) for crop simulation (National Climate Data Center, 2014). As an alternative to the use of GridWD or GenWD, here we present a protocol that utilizes 3 years of observed T_{max} and T_{min} data, combined with long-term GridWD of solar radiation and precipitation, to generate a long-term daily weather data set suitable for simulation of crop yields (hereafter called 'propagated' weather data [PWD]). The purpose of this paper is to evaluate how simulated yields compare when using PWD *versus* (i) OWD, (ii) GridWD and (iii) GenWD. In the present paper, the comparison was made across 18 sites, located in four continents (Europe, Asia, America, and Africa), for which long-term, high-quality daily OWD were available. Simulated crops include three major cereals (maize, rice and wheat), each simulated with well-validated crop models and based on site-specific soil properties and crop management to ensure agronomic relevance.

Table 1
Studies that used gridded or generated weather data for agricultural research in Sub-Saharan Africa.

Database	References
Gridded weather data	
CRU[a]	(Fischer et al., 2002; Foley et al., 2005; Bondeau et al., 2007; Lobell, 2007; Lobell et al., 2008; Battisti and Naylor 2009; Licker et al., 2010; Folberth et al., 2012; Folberth et al., 2013)
NASA[b]	(Folberth et al., 2012; Arndt et al., 2012)
NCEP[c]	(Lobell and Asner 2003; Nemani et al., 2003; Bagley et al., 2012)
WorldClim[d]	(Thornton et al., 2009; Nelson et al., 2010; Claessens et al., 2012)
Other[e]	(Jones and Thornton, 2003; Stéphenne and Lambin, 2001; Lobell et al., 2008; Rowhani et al., 2011)
Weather generators	
MarkSim	(Mavromatis and Hansen, 2001; Jones and Thornton, 2003; Thornton et al., 2009; Claessens et al., 2012)
WGEN (WeatherMan)	(Mavromatis and Hansen, 2001; Li et al., 2005; Schuol et al., 2008)
ClimGen	(Abraha and Savage, 2006; Laux et al., 2010)

[a] Climate Research Unit (CRU) http://badc.nerc.ac.uk/data/cru/.
[b] National Aeronautics and Space Administration (NASA) http://power.larc.nasa.gov/.
[c] National Center for Environmental Prediction/Department of Energy (NCEP) http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html.
[d] WorldClim http://www.worldclim.org/.
[e] All future climate data as modeled by global climate models, distributed by the International Panel on Climate Change (IPCC) http://www.ipcc-data.org/.

2. Materials and methods

2.1. Evaluation of NASA gridded data

The Prediction of Worldwide Energy Resource (POWER) dataset from the National Aeronautics and Space Administration (NASA, 2012), hereafter called NASA, was selected as the GridWD source for use in this study because it is publically accessible, shows general acceptable agreement with ground data for incident solar radiation, and has been used in previous studies that have simulated crop yields (Bai et al., 2010 Van Wart et al., 2013a,b). The NASA dataset contains daily incident solar radiation, T_{max}, T_{min}, dew point temperature (T_{dew}), precipitation, wind speed, and relative