10-25-2016

The Purdue Agro-climatic (PAC) Dataset for The U.S. Corn Belt: Development and Initial Results

Xing Liu

Elin Jacobs

Anil Kumar

Larry Biehl

Jeff Andersen

See next page for additional authors

Authors

Xing Liu, Elin Jacobs, Anil Kumar, Larry Biehl, Jeff Andersen, and Dev Niyogi

Climate Risk Management 15 (2017) 61–72

Contents lists available at ScienceDirect

Climate Risk Management

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

The Purdue Agro-climatic (PAC) dataset for the U.S. Corn Belt: Development and initial results

Xing Liu [a], Elin Jacobs [b], Anil Kumar [c], Larry Biehl [d], Jeff Andresen [e], Dev Niyogi [a,f,*]

[a] Department of Agronomy, Crops, Soils, and Environmental Science, Purdue University, West Lafayette, IN 47906, USA
[b] Lyles School of Civil Engineering, Purdue University, West Lafayette, IN 47906, USA
[c] ESSIC, University of Maryland, College Park, MD 20740, USA
[d] Information Technology at Purdue, Purdue University, West Lafayette, IN 47906, USA
[e] Department of Geology, Michigan State University, East Lansing, MI 48824, USA
[f] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 18 April 2016
Revised 13 October 2016
Accepted 24 October 2016
Available online 25 October 2016

Keywords:
Agroclimatology
Crop resiliency
Agriculture meteorology
Reanalysis
Land Data Assimilation System
Crop models

ABSTRACT

This study is a result of a project titled "Useful to Usable (U2U): Transforming Climate Variability and Change Information for Cereal Crop Producers". This paper responds to the project goal to improve farm resiliency and profitability in the U.S. Corn Belt region by transforming existing meteorological dataset into usable knowledge and tools for the agricultural community.

A high-resolution agro-climatic dataset that covers the U.S. Corn Belt was built for the U2U project based on a Land Data Assimilation System (LDAS) framework. This data referred to as the Purdue Agro-climatic (PAC) dataset is a gridded, continuous dataset suitable for agroclimatic and crop model studies over the U.S. Corn Belt. The dataset was created at 4 km, sub-daily spatiotemporal resolution and covers the period of 1981–2014. The dataset includes a range of variables such as daily maximum/minimum temperature, solar radiation, rainfall, evapotranspiration (ET), multilevel soil moisture and soil temperatures. The data were compared to field measurements from Ameriflux and the Soil Climate Analysis Network (SCAN), and with coarser but widely used atmospheric regional reanalysis data products. Validations indicate an overall good agreement between this dataset and field measurements. The agreement is particularly high for radiation and temperature parameters and lesser for rainfall and soil moisture data. Despite the differences with observations, the data show improvements over the coarser resolution products and other available models and thus highlights the value of the dataset for agroclimatic and crop model studies.

This high-resolution dataset is available to the wider community, and can fill gaps in observed data records and increase accessibility for the agricultural sector, and for conducting variety of if-then assessments.

1. Introduction

Agriculture is highly dependent on weather and climate. The U2U (www. Agclimate4u.org) project aims to "transform climate variability and change information for cereal crop producers" for improving the resiliency and profitability of farms

* Corresponding author at: Department of Agronomy, 915 W. State Street, Purdue University, West Lafayette, IN 47907-2054, USA.
E-mail address: dniyogi@purdue.edu (D. Niyogi).

http://dx.doi.org/10.1016/j.crm.2016.10.005
2212-0963/

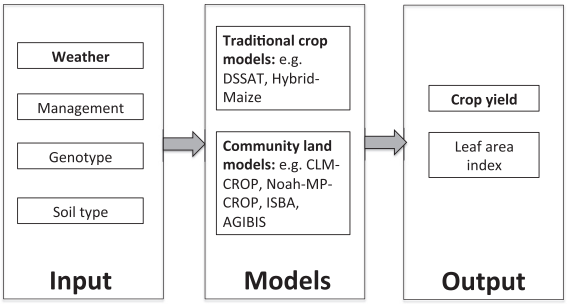

Fig. 1. Operational flow of large-scale crop modeling.

in the U.S. Corn Belt. This project seeks to deliver improved decision support tools, datasets and trainings. The U2U team is a diverse scientific group including climatologists, crop modelers, agronomists, economists, and social scientists (Prokopy et al., 2015). One of the objectives of climatologists and crop modelers group is to provide useful and usable dataset for users including crop modelers and producers (Niyogi and Andresen, 2011).

A majority of the agroclimatic assessments until now is based on point/field scale studies. Studies of food security under a changing climate and extreme weather, highlight an increasing demand for large spatial scale crop yield simulations (Hansen and Jones, 2000; Niyogi and Andresen, 2011; Rosenzweig et al., 2013; Takle et al., 2014; McDermid et al., 2015). As a result, a growing number of studies have been conducted on largescale crop simulations using traditional crop models (e.g., Rosenzweig et al., 2014; Elliott et al., 2014; Liu et al., 2015).

Fig. 1 summarizes the building blocks and the operational flow of such regional largescale simulations. Generally, the input data comprises of four groups: (i) weather (e.g. air temperature, solar radiation and precipitation), (ii) management practices (planting date, plant population and irrigation), (iii) plant genotype and (iv) regional soil texture and characteristics. These data are provided/needed at grid-by-grid basis across the study domain to the different crop simulation models. The models being run are either statistical models (e.g. Lobell et al., 2008) or traditional crop models, such as DSSAT (Jones et al., 2003), Hybrid-Maize (Yang et al., 2004), or part of land modeling system such as CLM-Crop (Drewniak et al., 2013), Noah-MP-Crop (Liu et al., 2016), ISBA (Garrigues et al., 2015), Agro-IBIS (Kucharik, 2003), ISAM (Song et al., 2013). The typical output of interest from these models are the crop yield, leaf area index, and evapotranspiration. The availability and usability of current input datasets however, are inadequate to fill the increasing demand for high spatiotemporal resolution regional crop simulations (Rosenzweig et al., 2013).

In this paper, we focus on one particular aspect of the data needs, those related to the weather input dataset. The regional agro-meteorological applications are often constrained by the spatially discontinuous meteorological data from regular weather stations. Further, the application of crop models is often limited by lack of hydro-meteorological input data, such as solar radiation, soil moisture and evaporation/transpiration. These variables are not routinely available from weather stations except for specific experimental field programs. The representation of spatial heterogeneity of weather and climate information is important for regional crop modeling (Doering, 2002; Niyogi et al., 2015). As a result, most models are run with default values or approximations based on empirical rules, and highlight the need for a high-resolution spatial, agro-climate dataset.

The climate community widely relies on reanalysis datasets that blend observations with detailed models in creating the gridded products (e.g. Kalnay et al., 1996; Mesinger et al., 2004). These reanalysis products are available as a scientific resource to the atmospheric community for a wide range of applications, and have also been a source of meteorological input for crop models studies. These datasets while suitable for large scale dynamical studies are generally too voluminous to store locally and too coarse for regional scale crop studies. Further, these data are not easy to use or work with for lay users (see for e.g. Table 1). Additionally, it is also difficult to extract the necessary data that is needed as an input for regional studies. Hence, an outstanding issue has been: how do we make these datasets useable for the broader agroclimate community, and crop modelers, more specifically?

Table 1
Examples of current reanalysis datasets (Including PAC dataset in this study).

Dataset	Time period	Highest Temporal resolution	Spatial Coverage	Typical Spatial resolution (Approximately)	Reference
NARR	1979–2015	3 h	North America	32 km	Mesinger et al. (2004)
MERRA-2	1980–present	3 h	Global	50 km	Rienecker et al. (2011)
NLDAS-2	1979–present	Hourly	North America	12 km	Mitchell et al. (2004)
AgMERRA	1980–2010	Daily	Global	27 km	Ruane et al. (2015)
Daymet	1980–2015	Daily	North America	1 km	Thornton et al. (2016)
PAC	1980–2014	Sub-daily	U.S. Corn Belt	4 km	This study

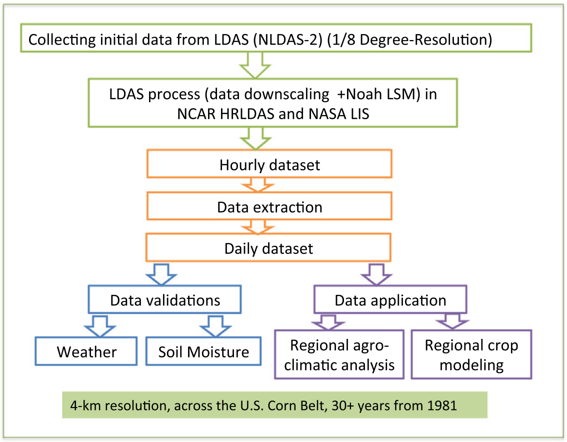

Fig. 2. Methodology flow chart for generating the PAC dataset.

This paper presents and builds on an approach that uses the Land Data Assimilation System (LDAS, ldas.gsfc.nasa.gov/) framework to create a high-resolution (4-km) agro-meteorological dataset: the Purdue Agro-climatic (PAC) dataset, to integrate weather and climate data suitable for crop-climate studies. Developing such a high-resolution dataset is expected to provide better access to tools that are needed for regional agricultural/climatic impact assessments and decision support studies.

Section 2 describes the process of developing the PAC dataset. Section 3 provides validations of this dataset with in situ meteorological data, along with the comparison with existing reanalysis based solar radiation and model generated solar radiation. Section 4 focuses primarily on the validations of soil moisture and soil temperature.

2. Datasets

The overall procedure is summarized in Fig. 2 and described further in this section.

At the heart of the dataset generation is a Noah land surface model (LSM) based Land Data Assimilation System (LDAS) framework. This system is used for downscaling and simulating surface hydrological parameters. The Noah LSM is a widely-used community model. It was developed on the concept of diurnally dependent Penman-based potential evaporation approach (Mahrt and Ek, 1984), the multilayer soil model (Mahrt and Pan, 1984), and a canopy transpiration model (Pan and Mahrt, 1987). Chen et al. (1996) extended this model by including the canopy resistance approach and Ek et al. (2003) added the formulation of bare soil. A large number of academic and operational research community users have developed this model further and is considered as a major component of the land/boundary layer atmospheric models, for both weather, hydrometeorology, and regional climate studies (Niu et al., 2011).

Originally, Noah LSM was developed to provide the land state for the NOAA/NCEP mesoscale Eta model (Betts et al., 1997; Chen et al., 1997; Ek et al., 2003). It has been included in LDAS, coupled with the Weather Research and Forecasting (WRF) regional atmospheric model. The Noah LDAS frameworks adopted in this study is based on the NCAR High Resolution LDAS (HRLDAS, Chen et al., 2007) and NASA Land Information System (LIS, Kumar et al., 2006).

In running the LDAS, the initial task was to compile different meteorological data into the NLDAS-2 (32-km resolution analysis). The NLDAS-2 uses bias-corrected GOES satellite-based downward shortwave radiation data, and precipitation data is mainly derived from hourly Doppler Stage II radar precipitation data (Mitchell et al., 2004). Additionally, land-surface initialization data (e.g., soil temperature, soil moisture, and canopy water content) were obtained from EDAS (Eta Data Assimilation System, Rogers et al., 1996). These were extracted to obtain different parameters separately into Grib files. A look up table as used in Noah/WRF was used to define the model land use/cover properties, terrain, soil texture, and monthly green vegetation fraction for Noah. The land-use input is based on 30-s U.S. Geological Survey (USGS) 24 categories. Terrain height is based on USGS-derived 30-s topographical height data, soil texture is based on the U.S. STATSGO soil map, and green vegetation fraction is based on monthly satellite-derived green vegetation fraction.

The next task was to downscale the raw meteorological data from 1/8 degree spatial resolution to 4-km grid spacing by running in a LDAS mode. This provides the foundation for high-resolution meteorological data that is integrated every hour and used for initializing landsurface conditions in the model at the start of each calendar year. The "input" data across the U.S. Corn Belt contain a total of 222,070 grids. The parameters included in each grid are listed in Supplementary Table S1. In this research, the hourly 4-km resolution meteorological data were grouped as "Database 1".

The 4-km resolution meteorological data was then used to drive the Noah LSM in a LDAS mode to simulate the soil conditions (e.g., soil moisture, soil temperature), ET (evapotranspiration), etc. During this process, Noah LSM simulates the surface conditions at a more detailed representation of topography, land cover, soil texture and vegetation type, obtained from

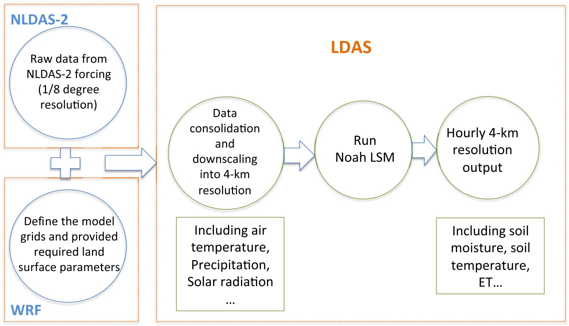

Fig. 3. The LDAS process flow.

Table 2
Variables included in the PAC daily dataset.

Name	Unit	Description
Tmax	°C	Daily maximum temperature at 2 m
Tmin	°C	Daily minimum temperature at 2 m
SR	MJ m{-2}	Daily solar radiation
Prep	mm	Daily precipitation
Soil_M	m{3} m{−3}	Daily averaged soil moisture (At 4 layer: 10 cm, 40 cm, 1 m, 2 m)
Soil_T	°C	Daily averaged soil temperature (At 4 layer: 10 cm, 40 cm, 1 m, 2 m)
ET	mm	Daily evapotranspiration

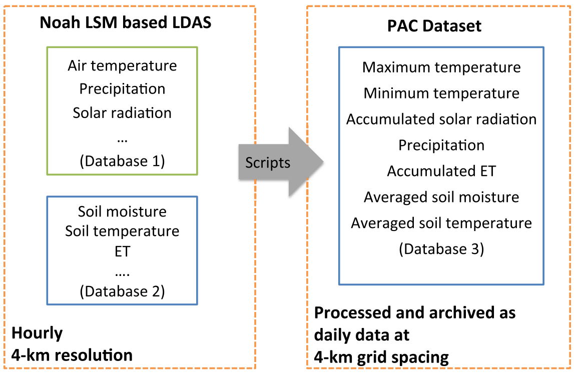

Fig. 4. Building the Purdue agro-meteorological dataset (PAC) from LDAS framework.

the high-resolution land cover information. The land model requires a "spin-up" period to account for hydro-dynamic balance. The "spin-up" time for Noah LSM typically requires few months (Chen et al., 2007; Charusombat et al., 2012). In this work, the spin-up was taken conservatively as 24 months (January 1979 to December 1980).

The output from LDAS is at a hourly and 4-km resolution for each grid. The output parameters generated for each grid are listed in Table S2. Fig. 3 presents the overall process of running the data processing and LDAS framework. The hourly 4-km resolution output data are grouped as "Database 2".

The objective of building PAC is to provide data that can be "useful and useable" for crop models and other agronomic decision tools. The minimum requirements of meteorological inputs for crop models (e.g., the Hybrid Maize model, Yang et al., 2004) include daily minimum temperature, daily maximum temperature, total solar radiation, and total precipitation. Therefore, to meet these needs, data extraction from the hourly database into daily data was necessary. A NCAR Command Language (NCL, Brown et al., 2012) script was developed and a module installed for data extraction. Careful attention had to be paid for ensuring data veracity while redoing the data file structures. For some variables, unit conversions were also necessary to make the data more usable (e.g. kg/m^2 of soil moisture to mm or m^3/m^3 etc).

The data extraction from hourly to daily was applied for different variables such as air temperature, precipitation, solar radiation, soil moisture and soil temperature. "Database 3", the PAC daily dataset was then compiled using these daily data (Table 2 and Fig. 4). A sample spatial plot for the maximum and minimum air temperature, daily precipitation, and daily solar radiation in shown in Fig. 5.

Since the domain covers different time zones, the data are stored in Universal Time Coordinate (UTC) system. If the daily meteorological data will be downscaled at local time, additional bias might be introduced due to different time zones. Here the data has not been corrected to local time because: (i) the research domain crosses three different time zones (Eastern Time, Central Time, and Mountain Time Zones); (ii) It is expected that daily maximum and minimum temperature are not significantly influenced by the time zone gap especially since they are developed from reviewing the hourly data. For example, in UTC, a day is defined from 00:00 to 00:00, while converted to the Eastern Time Zone the local time will be from previous day's 19:00 to current day 19:00. The daily maximum and minimum temperature usually occur during this time period. We also need to highlight that, the PAC dataset we presented here focuses on daily data, but the PAC framework can provide data at varying time-scales, from hourly to daily, so we call it as a "sub-daily" dataset.

3. Meteorological data validations

To validate the agro-meteorological database, 30-years (1981–2010) of observed temperature data for 18 counties (Fig. 6) were processed from the National Centers for Environmental Information (NCEI); solar radiation data for Bondville, IL were collected from Ameriflux (from 1997–2007) along with soil temperature/moisture data from different Ameiflux and SCAN sites. County-level yields were obtained from National Agricultural Statistics Service (NASS) annual survey, and are available as part of the broader dataset.

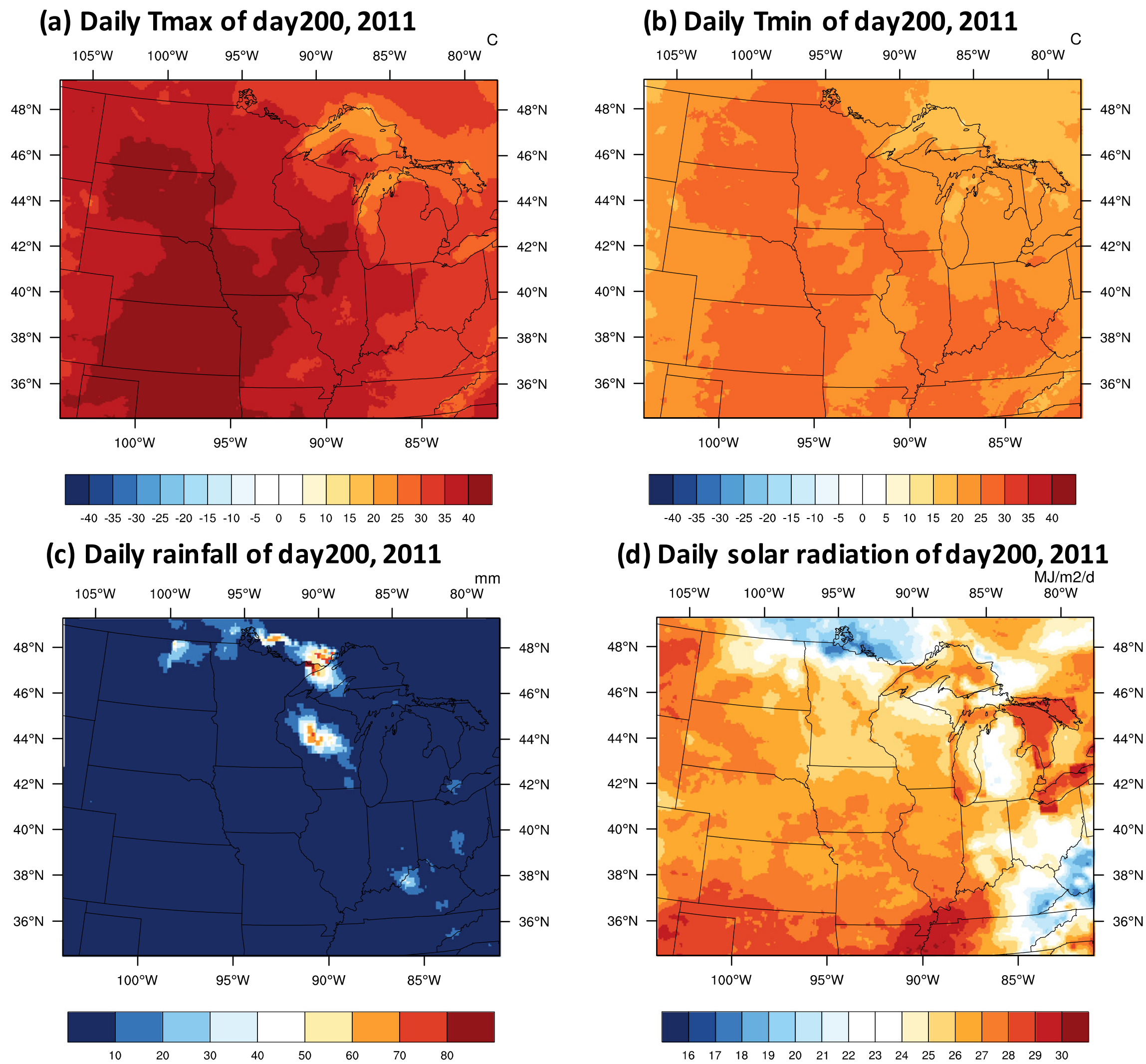

Fig. 5. Sample images of agrometeorological data from PAC: (a) Daily maximum temperature; (b) Daily minimum temperature; (c) Daily accumulated precipitation and (d) Daily solar radiation, for day 200 (i.e. 19 July) in 2011.

3.1. Maximum, minimum temperature and precipitation

The PAC dataset were compared with site daily observations, and the coefficient of determination (R^2) values are summarized in Table 3. The results indicate that the PAC daily maximum and minimum temperature have good agreement with the observations ($R^2 = 0.97$, for both maximum and minimum temperature). Since the data sample size is relatively large (~11,000 point for each site), Fig. 7 only shows the scatter plots for Johnson County, IA in 2001 as an example. For precipitation, the averaged R^2 is much lower and is 0.70. This is not surprising considering the rainfall can have both spatial and temporal errors (and also possibly due to a mismatch in the observed versus modeled day). Further, although the spatial resolution in PAC is 4-km, it is still difficult for reanalysis data to capture the spatial pattern and total amount of rainfall for a specific site. We also compared the PAC dataset with daily observations for growing season (April to October) only, the results are similar as the whole-year analysis, detailed results can be found in Supplementary Table S3.

3.2. Solar radiation

As mentioned before, crop models are often constrained by the lack of solar radiation data. The lack of data means, models have to rely on empirical approximations (Grant et al., 2004) or use data from synthetic weather generators such as

Fig. 6. Validation study domain and sites.

Table 3
Coefficient of Determination (R^2), Root-Mean-Square deviation (RMSE) and Bias between in situ daily observations and PAC reanalysis data at 18 sites for 30 years (1981–2010).

County	Tmax			Tmin			Precip		
	R^2	RMSE	Bias	R^2	RMSE	Bias	R^2	RMSE	Bias
Johnson, IA	0.98	2.65	0.99	0.98	2.52	−0.36	0.89	3.66	0.11
Winnebago, IA	0.97	3.26	−0.12	0.97	3.18	−1.63	0.70	5.56	−0.03
DeKalb, IL	0.96	3.25	0.31	0.97	3.17	−1.81	0.71	5.55	0.05
Douglass, IL	0.97	3.07	0.62	0.97	3.05	−1.67	0.70	5.99	−0.03
Huntington, IN	0.96	3.41	0.54	0.97	3.63	−2.4	0.63	5.86	0.05
Jasper, IN	0.96	3.33	0.01	0.97	2.95	−1.7	0.64	6.09	0.07
Shawnees, KS	0.97	2.97	−0.29	0.97	3.23	−1.92	0.74	5.59	0.08
Olmstead, MN	0.98	2.71	−0.29	0.98	2.99	−1.36	0.75	4.76	0.02
Renville, MN	0.97	3.69	−0.21	0.97	3.43	−1.97	0.69	4.66	−0.11
Adair, MO	0.97	3.07	−0.20	0.97	2.94	1.44	0.75	5.88	0.03
New Madrid, MO	0.94	3.40	−0.39	0.96	3.09	−1.66	0.66	7.73	0.04
Platte, NE	0.96	3.52	−0.67	0.97	3.18	−1.40	0.78	4.29	−0.02
Union, OH	0.97	2.76	0.71	0.97	2.93	−1.69	0.66	5.32	0.14
Rock, WI	0.96	3.44	0.23	0.97	3.01	−1.2	0.64	5.98	0.09
Sauk, WI	0.95	3.49	0.98	0.94	4.28	−2.5	0.60	5.90	0.06
Grand Forks, ND	0.98	3.53	−0.3	0.97	3.80	−1.57	0.74	3.71	0.04
Lucas, OH	0.98	2.53	0.46	0.96	3.11	−1.58	0.76	4.21	−0.01
Brookings, SD	0.97	3.93	−1.20	0.97	4.17	−3.13	0.71	4.29	0.02
Average	**0.97**	**3.22**	**0.07**	**0.97**	**3.26**	**−1.56**	**0.70**	**5.28**	**0.03**

WeatherAid (Yang et al., 2005). PAC provides daily solar radiation data, which can be used by not only crop models, but also other agronomic decision tools. The solar radiation data from PAC, which is based on satellite product (Mitchell et al., 2004), was compared with the observed solar radiation data from Bondville, IL, Ameriflux site. The validation results (Fig. 8) indicate a good fit with the observations ($R^2 = 0.81$). The solar radiation values from PAC were also compared against the weather generator, and the R^2 between generated solar radiation and measured observations is 0.67 (Fig. 9), results from Bondville site suggest the solar radiation data from PAC are potentially better than the solar radiation values generated by the weather generator. We also validated the daily solar radiation with another Ameriflux site: Mead, NE for year 2005, The R^2 is 0.69. In this study, due to the limitations of observations and data accessibility, we only presented results from two sites. More validation sites will be needed in the future studies that focus on solar radiation.

4. Soil moisture and soil temperature analysis

Soil moisture and soil temperature are important components of land-atmosphere interactions and critical variables in agrometeorology and crop production systems (Ochsner et al., 2013). Climate change and associated feedbacks in soil temperature and soil moisture are expected to affect agricultural systems with effects on crop productivity, crop variety, and planting and harvest times (Lobell et al., 2014). As mentioned in Section 1, hydroclimatic reanalysis products including soil moisture and temperature are available at coarse resolutions, and as a result not aligned with land surface model or crop model interfaces. In addition, while some in situ datasets for soil moisture and soil temperature measurements at the point scale are available, the quality of the datasets and record lengths vary. To validate the soil moisture and soil temperature estimates generated by the LDAS/Noah LSM, we compared (i) point observations from ten sites with corresponding model

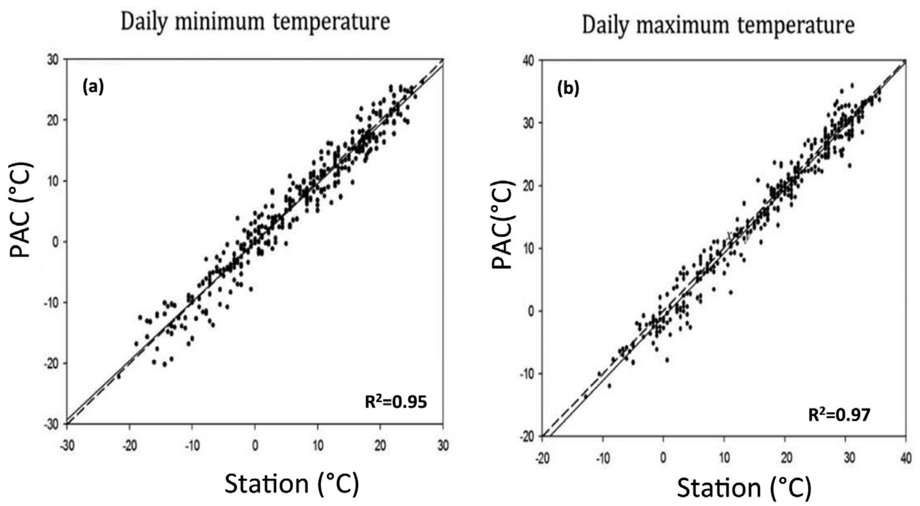

Fig. 7. (a) Minimum temperature, and (b) Maximum temperature for PAC dataset versus site observations for Johnson County, IA (2001).

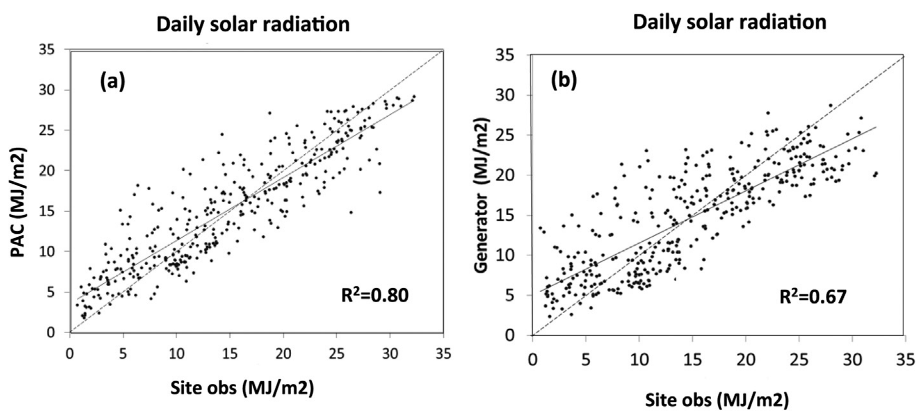

Fig. 8. (a) Daily solar radiation from PAC (grid) compared to the observations and (b) daily solar radiation from Weather generator (WeatherAid) vs. site observations, both plots are for Bondville, IL, 2001.

grids, and (ii) spatial representation of soil moisture with coarser resolution reanalysis products at the regional scale. A more comprehensive evaluation and application of the soil moisture data is reported in Niyogi et al. (in review) and the dissertation of Jacobs (2016). We focused here on the time period of interest, i.e. the growing season broadly defined as April through October consistent with other studies in the region (e.g. Kellner and Niyogi, 2015).

4.1. Point scale validation

Volumetric soil moisture and soil temperature observations from four USDA-NRCS Soil Climate Analysis Network (SCAN; Schaefer et al., 2007) and six Ameriflux (http://ameriflux.ornl.gov/) sites (Table 4) were analyzed. The sites were chosen on the basis of geographical distribution throughout the study domain. Results are compared with corresponding model grids at a monthly time step. Note that the data record lengths vary by site. Because observations of deeper soil layers were lacking or limited, the focus is on the topsoil layer (0–10 cm). Soil temperature estimates compared well with observations at all ten sites, with R^2 values generally greater than 0.90 (Table 5, Fig. 9). A few sites (e.g. Ames/Mandan/Johnson Farm) show larger

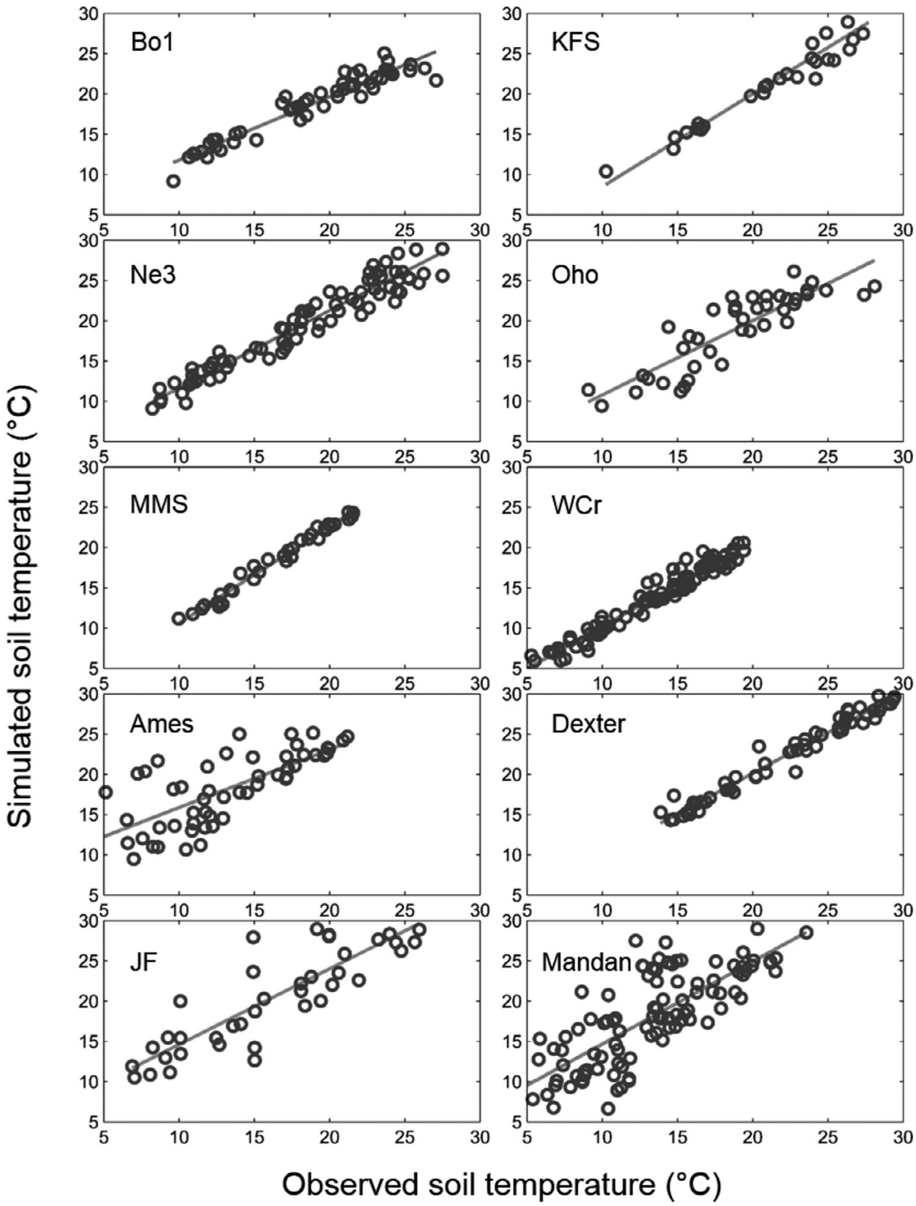

Fig. 9. Comparison of observed and simulated (PAC) top layer soil (0–10 cm) temperature for the growing season (circles), with linear regression fits (lines) at ten sites within the PAC domain.

Table 4
Observation sites used for comparison with the PAC data. Additional site information can be found at http://ameriflux.ornl.gov/ and http://www.wcc.nrcs.usda.gov/scan/. Note that the dates indicate the total length of the data record for all variables at each site. Soil moisture and soil temperature records may not be available for the entire record period.

Name	ID	State	Lat	Lon	LULC	Soil texture	From	To	Network
Bondville	Bo1	IL	40.006	−88.2904	Cropland	Silt loam	8/25/1996	11/4/2008	Ameriflux
Kansas Field Station	KFS	KS	39.056	−95.1907	Grasslands	Silt loam	6/16/2007	12/31/2012	Ameriflux
Mead rainfed	Ne3	NE	41.18	−96.4396	Cropland	Silt clay loam	5/25/2001	12/31/2012	Ameriflux
Ohio Oak Openings	Oho	OH	41.555	−83.8438	Deciduous broadleaf forest	Sand	1/1/2004	12/31/2011	Ameriflux
Morgan Monroe State Forest	MMS	IN	39.323	−86.4131	Deciduous broadleaf forest	Clay loam	1/1/1998	12/31/2010	Ameriflux
Willow Creek[1]	WCr	WI	45.806	−90.0798	Deciduous broadleaf forest	Sandy loam	1/1/1998	12/31/2012	Ameriflux
Ames	2031	IA	42.02	−93.73	Cropland	Clay loam	9/19/2001	12/31/2011	SCAN
Dexter	2048	MO	39.78	−89.93	Cropland	Silt loam	1/9/2001	12/31/2012	SCAN
Johnson Farm	2111	NE	40.37	−101.72	Cropland	Silt clay loam	10/1/2005	12/31/2012	SCAN
Mandan	2020	ND	46.77	−100.92	Grassland	Silt loam	1/1/1997	12/31/2012	SCAN

[1] The data for WCr are described in Cook et al. (2004)

Table 5
Coefficients of determination (R^2) for linear regression fits to growing season volumetric soil water content [$m^3\ m^{-3}$] and soil temperature [°C] of PAC product to in situ observations.

	Vol. soil-water content	Soil temperature
Site	R^2	R^2
Bo1	0.60	0.92
KFS	0.62	0.93
Ne3	0.59	0.93
Oho	0.39	0.76
MMS	0.77	0.98
WCr	0.24	0.96
Ames	0.50	0.59
Dexter	0.45	0.96
JF	0.70	0.72
Mandan	0.57	0.68

variability between observed and modeled soil temperature. This could be due to the monitoring equipment used at the sites, microclimatic differences that are averaged out over the larger grid scale, and model error. For example, quality control of observed soil moisture data sets is variable, Recently, efforts to automate the quality control of network data have been undertaken (Xia et al., 2015a). Further analysis is underway to determine the cause of these discrepancies (Jacobs, 2016; Niyogi et al., in review). It is worth to note that the sites with the largest temperature deviations perhaps coincidentally belong to the SCAN network.

Point scale comparisons of observed and modeled soil moisture show that there are larger deviations as compared to soil temperature (Fig. 10, Table 5). This is to be expected due to the more complex nature of the soil hydrologic processes and related soil properties. Yet, the R^2 are generally above 0.50 and in some cases above 0.70. The model over the sites with soils containing a large fraction of sand typically performed worse than finer soil types (i.e. Ohio Oak Forest, Willow Creek). Also the model versus observed values for winter season show large discrepancies. After contacting the site scientists, these discrepancies were narrowed down to the high uncertainty in the measurement protocols for winter months and, also due to the error that persist in the model for snow cover period (Barlage et al., 2015). There is a large number of studies focusing on validation of modeled soil moisture (e.g. Koster et al., 2009; Xia et al., 2015b; Coopersmith et al., 2016). Volumetric soil moisture is variable over short distances due to diverse soil types, land-cover, and topographic changes (Xia et al., 2015c; Coopersmith et al., 2016). Most LSMs are run at a relatively coarse grid scale (~1–100 km) and to simplify the diversity in surface and subsurface properties each grid cell represents the dominant soil type, vegetation type, and topographic condition over each model grid cell. Because the spatial variability within a grid cell is not fully represented in the LSM, disparity between the model output and in situ observations are common. Soil moisture sensors are generally geographically sparse and, depending on region, only one site may be available within a model grid cell which makes it difficult to fully analyze the reason behind biases between observations and models in terms of spatial variability vs. model limitations (Xia et al., 2015b). Others claim that simulated soil moisture should not be treated as equivalent to observed soil moisture at all, but rather viewed as a wetness index used to balance water losses through evapotranspiration and runoff (Koster et al., 2009). Again, the reasons for discrepancies between observations and PAC estimates are not clear, but are likely due to the point to grid scale differences, soil hydrology model parameterization and the soil information used as model input (see e.g. Chen and Dudhia, 2001 for details).

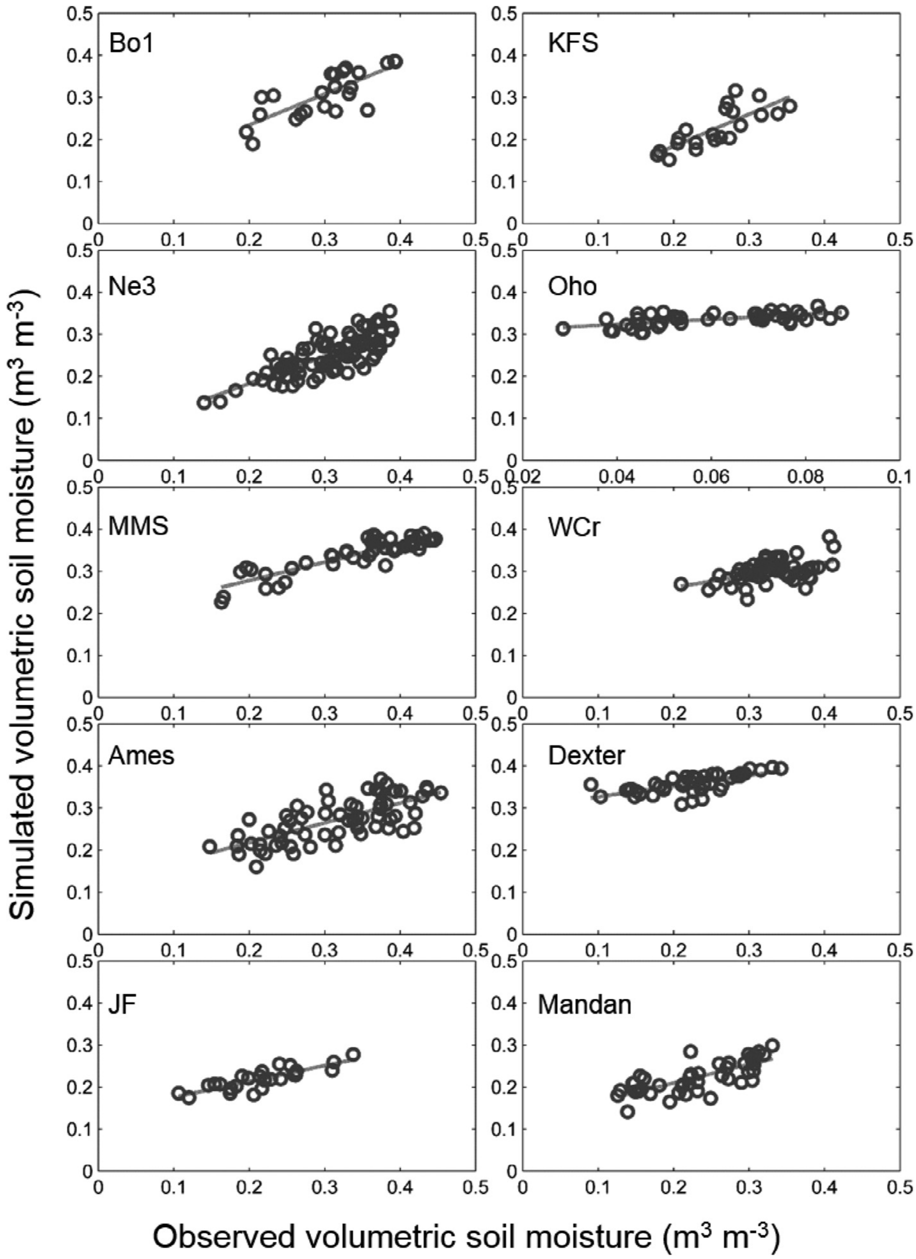

Fig. 10. Same as Fig. 9 but for volumetric soil moisture content. Note the difference in x-axis scale for Oho.

5. Conclusion

The goal of building this high resolution agro-meteorological PAC dataset is to bring available meteorological reanalysis information to usable agronomic applications, such as crop models. This goal was achieved by using a Land Data Assimilation System (LDAS) framework, and hydrodynamically downscaling data from 32-km into 4-km grid spacing in Noah LSM. The LDAS output based on the NCAR HRLDAS and NASA LIS recomputed the surface energy and water balance at the new resolution and corresponding land cover, soil texture, and topography; by processing the LDAS field hourly, regional agroclimatic dataset was created. To help most agronomic applications, a daily database of 30+ years (1981–2014) was built, and includes variables such as maximum and minimum air temperature, solar radiation, precipitation, surface ET, and soil moisture and soil temperature at different depths. Results of the initial evaluation undertaken indicate that the variables in the agro-meteorological database show good agreement with in situ data and other popular reanalysis datasets. Data from PAC also showed a better fit with observations especially for solar radiation particularly when compared with that from a weather generator output. These results are encouraging and provide confidence to apply this high-resolution agro-meteorological database in agronomic applications. The availability of the PAC dataset helps provide better access to agroclimatic dataset in term of data resolution, quality and data continuity. These data are expected to help investigations seeking to study the influence of climate on crop growth at the regional scales over the U.S. Corn Belt (e.g. Liu et al., 2016).

Acknowledgments

This research is part of USDA National Institute of Food and Agriculture project titled "Useful to Usable (U2U): Transforming Climate Variability and Change Information for Cereal Crop Producers," and is supported by Competitive Grant no. 2011-68002-30220, and Hatch project 1007699. Study also benefit from NSF CAREER (AGS-0847472), NOAA/NAS/AFWA Developmental Test Center project with NCAR, Texas A&M project on drought trigger (competitive grant no. 2011-67019-20042), and competitive grant no. 2015-67023-23109. The PAC dataset in archived and distributed via U2U (www.agclimate4u.org) data portal, and will be available for accessing publicly.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.crm.2016.10.005.

References

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.L., Niu, G.Y., 2015. The effect of groundwater interaction in North American regional climate simulations with WRF/Noah-MP. Clim. Change 129 (3–4), 485–498.
Betts, A.K., Chen, F., Mitchell, K.E., Janjic, Z.I., 1997. Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data. Mon. Weather Rev. 125 (11), 2896–2916.
Brown, D., Brownrigg, R., Haley, M., Huang, W., 2012. The NCAR Command Language (NCL)(version 6.0. 0). UCAR/NCAR Computational and Information Systems Laboratory, Boulder, CO.
Charusombat, U., Niyogi, D., Garrigues, S., Olioso, A., Marloie, O., Barlage, M., Chen, F., Ek, M., Wang, X., Wu, Z., 2012. Noah-GEM and Land Data Assimilation System (LDAS) based downscaling of global reanalysis surface fields: Evaluations using observations from a CarboEurope agricultural site. Comput. Electron. Agric. 86, 55–74.
Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.L., Koren, V., Duan, Q.Y., Ek, M., Betts, A., 1996. Modeling of land surface evaporation by four schemes and comparison with FIFE observations. J. Geophys. Res. 101 (D3), 7251–7268.
Chen, F., Janjić, Z., Mitchell, K., 1997. Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. Bound.-Layer Meteorol. 85 (3), 391–421.
Chen, F., Manning, K.W., LeMone, M.A., Trier, S.B., Alfieri, J.G., Roberts, R., Tewari, M., Niyogi, D., Horst, T.W., Oncley, S.P., Basara, J.B., Blanken, P.D., 2007. Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. J. Appl. Meteorol. Climatol. 46 (6), 694–713.
Chen, F., Dudhia, J., 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Weather Rev. 129 (4), 569–585.
Cook, B.D., Davis, K.J., Wang, W., Desai, A., Berger, B.W., Teclaw, R.M., Martin, J.G., Bolstad, P.V., Bakwin, P.S., Yi, C., Heilman, W., 2004. Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin. Agric. Forest Meteorol. 126 (3), 271–295.
Coopersmith, E.J., Cosh, M.H., Bell, J.E., Kelly, V., Hall, M., Palecki, M.A., Temimi, M., 2016. Deploying temporary networks for upscaling of sparse network stations. Int. J. Appl. Earth Obs. Geoinf. 52, 433–444. http://dx.doi.org/10.1016/j.jag.2016.07.013.
Doering III, O.C., 2002. Effects of Climate Change and Variability on Agricultural Production Systems. Springer Science & Business Media. 278pp.
Drewniak, B., Song, J., Prell, J., Kotamarthi, V.R., Jacob, R., 2013. Modeling agriculture in the community land model. Geosci. Model Dev. 6 (2), 495–515.
Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. Atmos. 108 (D22), 8851.
Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K.J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R.C., 2014. The Global Gridded Crop Model intercomparison: data and modeling protocols for Phase 1 (v1. 0). Geosci. Model Dev. Discuss. 7 (4), 4383–4427.
Garrigues, S., Olioso, A., Carrer, D., Decharme, B., Calvet, J.C., Martin, E., Moulin, S., Marloie, O., 2015. Impact of climate, vegetation, soil and crop management variables on multi-year ISBA-A-gs simulations of evapotranspiration over a Mediterranean crop site. Geosci. Model Dev. 8 (10), 3033–3053.
Grant, R.H., Hollinger, S.E., Hubbard, K.G., Hoogenboom, G., Vanderlip, R.L., 2004. Ability to predict daily solar radiation values from interpolated climate records for use in crop simulation models. Agric. For. Meteorol. 127 (1), 65–75.
Hansen, J.W., Jones, J.W., 2000. Scaling-up crop models for climate variability applications. Agric. Syst. 65 (1), 43–72.
Jacobs, E.M., 2016. Spatiotemporal Patterns of Hydroclimatic Drivers and Soil-Water Storage: Observations and Modeling Across Scales (Doctoral Dissertation). Purdue University, West Lafayette, IN. Available from Purdue University Library.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3), 235–265.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–471.
Kellner, O., Niyogi, D., 2015. Climate Variability and the US Corn Belt: ENSO and AO Episode-Dependent Hydroclimatic Feedbacks to Corn Production at Regional and Local Scales. Earth Interact. 19 (6), 1–32.
Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., Puma, M.J., 2009. On the nature of soil moisture in land surface models. J. Clim. 22 (16), 4322–4335.
Kucharik, C.J., 2003. Evaluation of a process-based agro-ecosystem model (Agro-IBIS) across the US corn belt: simulations of the interannual variability in maize yield. Earth Interact. 7 (14), 1–33.
Kumar, S.V., Peters-Lidard, C.D., Tian, Y., Houser, P.R., Geiger, J., Olden, S., Lighty, L., Eastman, J.L., Doty, B., Dirmeyer, P., Adams, J., 2006. Land information system: An interoperable framework for high resolution land surface modeling. Environ. Model. Software 21 (10), 1402–1415.
Liu, X., Andresen, J., Yang, H.S., Niyogi, D., 2015. Calibration and validation of the hybrid-maize crop model for regional analysis and application over the US Corn Belt. Earth Interact. 19, 1–16.
Liu, X., Chen, F., Barlage, M., Niyogi, D., 2016. (In revision): Noah-MP-Crop: introducing dynamic crop growth in the Noah-MP land-surface model. J. Geophys. Res. Atmos.
Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319 (5863), 607–610.

Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., Rejesus, R.M., Hammer, G.L., 2014. Greater sensitivity to drought accompanies maize yield increase in the US Midwest. Science 344 (6183), 516–519.

Pan, H.L., Mahrt, L., 1987. Interaction between soil hydrology and boundary-layer development. Bound.-Layer Meteorol. 38 (1–2), 185–202.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.K., Bloom, S., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. J. Clim. 24 (14), 3624–3648.

Rogers, E., Black, T.L., Deaven, D.G., DiMego, G.J., Zhao, Q., Baldwin, M., Junker, N.W., Lin, Y., 1996. Changes to the operational "early" Eta analysis/forecast system at the National Centers for Environmental Prediction. Weather Forecasting 11 (3), 391–413.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A. M., Schmid, E., Stehfesk, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. 111, 3268–3273.

Ruane, A.C., Goldberg, R., Chryssanthacopoulos, J., 2015. Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. Agric. For. Meteorol. 200, 233–248.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wei, Y., Devarakonda, R., Vose, R.S., Cook, R.B., 2016. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed October, 2016.

Mahrt, L., Ek, M., 1984. The influence of atmospheric stability on potential evaporation. J. Appl. Meteorol. 23, 222–234.

Mahrt, L., Pan, H., 1984. A two-layer model of soil hydrology. Bound.-Layer Meteorol. 29 (1), 1–20.

McDermid, S., Ruane, A., Hudson, N.I., Rosenzweig, C., Ahuja, L.R., Anapalli, S.S., 2015. The AgMIP coordinated climate-crop modeling project (C3MP): methods and protocols. In: Handbook of Climate Change and Agroecosystems: The Agricultural Model Intercomparison and Improvement Project (AgMIP)-Integrated Crop and Economic Assessments. World Scientific Publishing Co Ltd.

Mesinger, F., DiMego, G., Kalnay, E., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Mitchell, K., Rogers, E., Ek, M., Fan, Y., 2004. NCEP North American regional reanalysis. Am. Meteorol. Soc.

Mitchell, K.E., Lohmann, D., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., 2004. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J. Geophys. Res. Atmos. 109 (D7).

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res. Atmos. 116, D12.

Niyogi, D., Andresen, J., 2011. Useful to Usable (U2U): transforming climate variability and change information for cereal crop producers. 2011 Fall Meeting, San Francisco, CA, Amer. Geophys. Union. Abstract GC13A-0960.

Niyogi, D., Liu, X., Andresen, J., Song, Y., Jain, A.K., Kellner, O., Takle, E.S., Doering, O.C., 2015. Crop models capture the impacts of climate variability on corn yield. Geophys. Res. Lett. 42, 3356–3363. http://dx.doi.org/10.1002/2015GL063841.

Niyogi, D., Jacobs, E.M., Liu, X., Kumar, A., Biehl, L., Rao, P.S., 2016. (In review): Long-term high resolution hydroclimatic dataset for the U.S. Midwest. Earth Interact.

Ochsner, T.E., Cosh, M.H., Cuenca, R.H., Dorigo, W.A., Draper, C.S., Hagimoto, Y., Kerr, Y.H., Njoku, E.G., Small, E.E., Zreda, M., 2013. State of the art in large-scale soil moisture monitoring. Soil Sci. Soc. Am. J. 77 (6), 1888–1919.

Prokopy, L.S., Hart, C.E., Massey, R., Widhalm, M., Klink, J., Andresen, J., Angel, J., Blewett, T., Doering, O.C., Elmore, R., Gramig, B.M., Guinan, P., Hall, B.L., Jain, A., Knuton, C., Lemos, M.C., Morton, L.W., Niyogi, D., Power, R., Shulski, M.D., Song, C.X., Takle, E.S., Todey, D., 2015. Using a team survey to improve team communication for enhanced delivery of agro-climate decision support tools. Agric. Syst. 138, 31–37.

Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The USDA natural resources conservation service soil climate analysis network (SCAN). J. Atmos. Oceanic Technol. 24 (12), 2073–2077.

Song, Y., Jain, A.K., McIsaac, G.F., 2013. Implementation of dynamic crop growth processes into a land surface model: evaluation of energy, water and carbon fluxes under corn and soybean rotation. Biogeosciences 10 (12), 8039–8066.

Takle, E.S., Anderson, C.J., Andresen, J., Angel, J., Elmore, R.W., Gramig, B.M., Guinan, P., Hilberg, S., Kluck, D., Massey, R., Niyogi, D., 2014. Climate forecasts for corn producer decision making. Earth Interact. 18 (5), 1–8.

Xia, Y., Ford, T.W., Wu, Y., Quiring, S.M., Ek, M.B., 2015a. Automated Quality Control of In Situ Soil Moisture from the North American Soil Moisture Database Using NLDAS-2 Products. J. Appl. Meteorol. Climatol. 54 (6), 1267–1282.

Xia, Y., Ek, M.B., Wu, Y., Ford, T., Quiring, S.M., 2015b. Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis. J. Hydrometeorol. 16 (5), 1962–1980.

Xia, Y., Ek, M.B., Wu, Y., Ford, T., Quiring, S.M., 2015c. Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part II: Impact of soil texture classification and vegetation type mismatches. J. Hydrometeorol. 16 (5), 1981–2000.

Yang, H.S., Dobermann, A., Lindquist, J.L., Walters, D.T., Arkebauer, T.J., Cassman, K.G., 2004. Hybrid-maize—a maize simulation model that combines two crop modeling approaches. Field Crops Res. 87 (2), 131–154.

Yang, H.S., Dobermann, A., Cassman, K.G., Walters, D.T., 2005. WeatherAid: A Software for Weather Data Management. University of Nebraska, Lincoln.