Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation

Alex C. Ruane [a],[*], Richard Goldberg [b], James Chryssanthacopoulos [b]

[a] NASA Goddard Institute for Space Studies, New York, NY, United States
[b] Columbia University Center for Climate Systems Research, New York, NY, United States

ABSTRACT

The AgMERRA and AgCFSR climate forcing datasets provide daily, high-resolution, continuous, meteorological series over the 1980–2010 period designed for applications examining the agricultural impacts of climate variability and climate change. These datasets combine daily resolution data from retrospective analyses (the Modern-Era Retrospective Analysis for Research and Applications, MERRA, and the Climate Forecast System Reanalysis, CFSR) with in situ and remotely-sensed observational datasets for temperature, precipitation, and solar radiation, leading to substantial reductions in bias in comparison to a network of 2324 agricultural-region stations from the Hadley Integrated Surface Dataset (HadISD). Results compare favorably against the original reanalyses as well as the leading climate forcing datasets (Princeton, WFD, WFD-EI, and GRASP), and AgMERRA distinguishes itself with substantially improved representation of daily precipitation distributions and extreme events owing to its use of the MERRA-Land dataset. These datasets also peg relative humidity to the maximum temperature time of day, allowing for more accurate representation of the diurnal cycle of near-surface moisture in agricultural models. AgMERRA and AgCFSR enable a number of ongoing investigations in the Agricultural Model Intercomparison and Improvement Project (AgMIP) and related research networks, and may be used to fill gaps in historical observations as well as a basis for the generation of future climate scenarios.

Published by Elsevier B.V.

1. Introduction

The Agricultural Model Intercomparison and Improvement Project (AgMIP; Rosenzweig et al., 2013a) is conducting a wide range of climate-impacts-oriented activities focusing on crop and livestock models at the local level (e.g., Asseng et al., 2013; Singels et al., 2013; Bassu et al., 2014; Li et al., 2014; Ruane et al., 2014b) and on a global grid (Rosenzweig et al., 2013b), regional assessments of food security (Rosenzweig et al., 2012), and global economic impacts (e.g., Nelson et al., 2013; von Lampe et al., 2014). Related regional research networks such as the Consultative Group on International Agricultural Research (CGIAR) Climate Change, Agriculture and Food Security (CCAFS) and MACSUR (Modeling European Agriculture with Climate Change for Food Security; Rötter et al., 2013) are dealing with similar tasks. Consistency and transparency in climate data and methods facilitate comparisons across regions or between models in each of these assessments, particularly when market linkages between regions are emphasized. In

particular, recent advances in porting agricultural models for parallel processing on high-performance computing has dramatically increased the demand for global climate datasets capable of driving global gridded crop models (Rosenzweig et al., 2013b). The historical period is of primary and urgent interest, as data from recent years may be used to calibrate models and serve as the basis for the development of future climate scenarios using different statistical methods (Wilby et al., 2004).

Here we describe the development of two new climate forcing datasets (AgMERRA and AgCFSR) designed to meet the needs of AgMIP and similar agricultural impacts assessments (White et al., 2011a). As opposed to strictly climatic datasets, particular consideration is given to agricultural areas and the climatic factors that crops are known to respond to, including biases in mean growing season temperature and precipitation, the seasonal cycle, interannual variability, the frequency and sequence of rainfall events, and the distribution of sub-seasonal extremes.

The root of all climate forcing datasets is the network of in situ meteorological observations maintained by meteorological agencies around the world. The density and quality of these stations varies widely through space and time, with the best coverage in developed countries and less reliable coverage in the Tropics and

Table 1
Overview of Climate Forcing Datasets, including the AgMERRA and AgCFSR datasets introduced here. Highest resolution is the resolution at which the data are archived and most finely distinguishable, although for some variables multiple grid boxes may be given the same value as the effective resolution is more coarse.

Climate Forcing Dataset	Reference	Time period	Highest resolution	Reanalysis basis (and resolution)	Monthly target for temperature and precipitation
Princeton	Sheffield et al. (2006)	1948–2008	$0.5° \times 0.5°$	Reanalysis-1 ($\sim 2°$)	CRU TS2.0, with corrections for high-latitude precipitation using GPCP and TRMM
WFD		1958–2001	$0.5° \times 0.5°$	ERA-40 ($1°$)	CRU TS2.1 and GPCCv4 versions
WFD-EI	Weedon et al. (2012)	1979–2009	$0.5° \times 0.5°$	ERA-Interim ($0.4°$)	CRU TS3.1 and GPCCv5/6 versions
GRASP	Iizumi et al. (2014)	1961–2010	$1.125° \times 1.125°$	JRA25 ($1.125°$) and ERA-40 ($2.5°$ version)	CRU TS3.10.01, time-constant correction factors derived from 1961 to 1990.
AgMERRA	This study	1980–2010	$0.25° \times 0.25°$	MERRA ($0.5° \times 0.67°$)	Blend of in situ (CRU TS3.1, GPCCv6, WM) and satellite (TRMM, CMORPH, PERSIANN) products
AgCFSR	This study	1980–2010	$0.25° \times 0.25°$	CFSR ($\sim 0.3°$)	Blend of in situ (CRU TS3.1, GPCCv6, WM) and satellite (TRMM, CMORPH, PERSIANN) products

Southern Hemisphere (Lorenz and Kunstmann, 2012). These data are also not always accessible and transparent as they may require high acquisition fees, restrictive limitations on use, or additional processing and quality control beyond the scope of many agricultural modelers. Several groups have collected these data and constructed harmonized, global gridded datasets at monthly resolution (New et al., 2002; Schneider et al., 2011; Willmott and Matsuura, 1995; Hijmans et al., 2005), however these require weather generators to synthesize daily resolution before they may be applied to crop models and are therefore likely to miss events that are important to the calibration and validation of agricultural models. Regional gridded observational networks have also been created (e.g., E-Obs in Europe, Haylock et al., 2008; APHRODITE in Asia, Yatagai et al., 2012; CPC US Unified Precipitation, Higgins et al., 2000), however many regions and variables are not covered by any such network and intercomparing sites between regions with different methodologies introduces inconsistencies.

The overall meteorological observational network is larger than just stations, as weather balloons and airborne instruments provide information about the upper atmosphere and satellite-based observations (particularly beginning in the late 1970s and including direct estimates of precipitation since the late 1990s) augment the entire network. The atmospheric modeling community has developed retrospective-analyses (reanalyses) that assimilate all available state observations into a physically-consistent atmospheric model that utilizes atmospheric structure and dynamics to estimate spatial and variable gaps in the observations. These reanalyses were designed for process studies, emphasizing atmospheric structure and circulation over some impacts-relevant variables. Flux variables, such as precipitation and radiation, are modeled rather than assimilated. Additionally, 2-m temperature, wind speed, and humidity measurements are not assimilated, as reanalyses rely instead on balloon (rawinsonde) networks to assimilate in the free atmosphere and then model boundary–layer profiles. The adherence to physical principles can lead to biases even at assimilated locations where limitations in model parameterizations or spatial resolution cannot be overcome.

In an effort to correct some of the most glaring shortcomings of the reanalyses, the land-surface hydrology community led the development of climate forcing datasets that adjust the reanalyses' daily time series to match the monthly gridded climate datasets. This can prevent full closure of the water and energy cycles, but maintains many of the most important properties for impacts assessment. Schwalm et al. (2014) found that hydrologic models

are quite sensitive to the selection of a climate forcing dataset in the US, but only recently has the same question been asked of the agricultural models (e.g., Ruane et al., 2014a; Iizumi et al., 2014) despite the fact that agricultural models do not have the benefit of aggregating potentially compensating errors across watersheds. Adam et al. (2006) note that many global gridded climate datasets are biased toward the populated areas where stations have been set up rather than the mountains surrounding these, for example. This bias may be problematic for hydrologic catchments, but likely benefits agricultural applications as farmlands tend to be in the valleys and plains that are overrepresented.

This paper presents two new climate forcing datasets developed for agricultural applications utilizing a newer generation of reanalyses that are not currently associated with any climate forcing dataset. These reanalyses' higher spatial resolution, improved model physics, and additional sources of assimilated data hold great potential for improved agroclimatic assessment. Section 2 describes the datasets used in the construction, calibration, and evaluation of the AgMERRA and AgCFSR climate forcing datasets. Section 3 details the specifications of these new datasets and provides the complete methodology for their generation. Section 4 compares AgMERRA and AgCFSR against observations, the original reanalyses that they are drawn from, and existing climate forcing datasets. Following a discussion of the datasets' strengths and weaknesses, we describe the potential for gap-filling applications. Finally, we provide conclusions and next steps in the development, extension, and application of climate forcing datasets for agricultural modeling.

2. Datasets

2.1. Climate datasets

2.1.1. Existing climate forcing datasets

Methodologies for the development of the AgMIP climate forcing datasets was motivated by similar climate forcing datasets developed for various applications in recent years (Table 1), with the hopes that that new datasets could provide dramatically improved sub-monthly weather characteristics and radiation data that would improve agricultural modeling. The Princeton Climate Forcing Dataset (Sheffield et al., 2006) was developed for hydrologic applications, deriving its daily time series from the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis-1 (Kalnay et al., 1996) and adjusting to match

Table 2

Summary of variable construction methodologies for the AgMERRA and AgCFSR climate forcing datasets. The effective resolution for temperature and radiation is higher than the 1/4° resolution of the climate forcing datasets if there is heavy reliance on a particular observational dataset, with the 1/4° resolution only coming from the interpolation of a more coarse value. DTR = Diurnal temperature range.

Variable (and units)	Effective resolution	AgMERRA construction summary	AgCFSR construction summary
Maximum and minimum temperature (°C)	0.5°	*Mean:* MERRA daily T_{max} and T_{min} values shifted by average monthly temperature correction from CRU and WM for each month in each year on 1/2° grid *DTR:* Adjusted to be 3/4 of the way between MERRA and CRU DTRs. Ensure that $T_{max} > T_{min}$	*Mean:* CFSR daily T_{max} and T_{min} values shifted by average monthly temperature correction from CRU and WM for each month in each year on 1/2° grid *DTR:* Adjusted to be equivalent to CRU DTR. Ensure that $T_{max} > T_{min}$
Precipitation (mm/day)	0.25°	*Wet days:* Average of MERRA-Land and CRU wet days for each month in each year *Mean:* MERRA-Land daily values multiplied by correction factor imposing mean of CRU, GPCC, and WM for each month and each year at 1/2 resolution. 1/4° detail imposed from average monthly spatial pattern drawn from ensemble of TRMM, CMORPH, and PERSIANN	*Wet days:* CRU wet days for each month in each year *Mean:* CFSR daily values multiplied by correction factor imposing mean of CRU, GPCC, and WM for each month and each year at 1/2 resolution. 1/4° detail imposed from average monthly spatial pattern drawn from ensemble of TRMM, CMORPH, and PERSIANN
Solar radiation (MJ/m^2/day)	1.0°	*07/1983–12/2007:* NASA/GEWEX SRB data linearly interpolated to 1/4° grid *01/1980–06/1983 and 01/2007–12/2010:* MERRA downward shortwave flux corrected using quantile-mapping and the statistics of SRB Beta distribution	*07/1983–12/2007:* NASA/GEWEX SRB data linearly interpolated to 1/4° grid *01/1980–06/1983 and 01/2007–12/2010:* CFSR downward shortwave flux corrected using quantile-mapping and the statistics of SRB Beta distribution
Relative humidity at the time of maximum temperature (%)	0.25°	Calculated from MERRA specific humidity, maximum temperature, and surface pressure and then linearly interpolated to 1/4° grid.	Calculated from CFSR specific humidity, maximum temperature, and surface pressure and then linearly interpolated to 1/4° grid
Wind speed (m/s)	0.25°	MERRA wind speeds linearly interpolated to 1/4° grid	Adjusted CFSR 10-m wind speeds to 2-m velocities and then linearly interpolated to 1/4° grid

CRU monthly temperature and precipitation totals. The Water and Global Change (WATCH) climate forcing dataset (WFD; Weedon et al., 2011) was also developed with a hydrologic focus, using the European Centre for Medium-range Weather Forecasting 40 year reanalysis (ERA-40; Uppala et al., 2005) and adjusting to match CRU monthly temperature and precipitation totals from CRU or GPCC. As improved and higher-resolution reanalyses have been developed to replace the ERA-40 and NCEP/NCAR Reanalysis-1, WATCH has also created a second climate forcing dataset (WFD-EI; Weedon et al., 2011) applying its methodology to the next generation ERA-Interim reanalysis (Dee et al., 2011). The GPCC corrected versions of the WATCH datasets are used in the evaluations below. Iizumi et al. (2014) have also recently created the Global Risk Assessment for the Stable Production of Food (GRASP) meteorological forcing dataset with an explicit agricultural focus, using a combination of the 25-year Japanese Reanalysis (JRA-25; Onogi et al., 2007) and the ERA-40 (in earlier years) and adjusting to match CRU monthly temperature and precipitation totals using time-constant correction factors derived from a comparison over the 1961–1990 period. Many of these products also systematically correct the number of rainy days, humidity, solar radiation, and wind speed (see Table 2 of Iizumi et al., 2014, for a review). The AgMIP climate forcing datasets build upon these established methods, adding improved datasets and features described below to produce new datasets that enable AgMIP and related agricultural applications.

2.1.2. Original reanalyses

NASA's Modern-Era Retrospective Analysis for Research and Applications (MERRA; Rienecker et al., 2011) forms the basis of the AgMERRA climate forcing dataset. MERRA was designed to cover the satellite era (post-1979) with a particular focus on the water cycle, and provides hourly output of surface meteorological fields on a 1/2° latitude by 2/3° longitude grid. AgMERRA also utilizes MERRA-Land (Reichle et al., 2011), a version with additional assimilation of the 1/2° × 1/2° Climate Prediction Center's Unified precipitation product (CPCU; Chen et al., 2008) from 1980 to 2005 and the CPC's real-time product from 2006 to 2010 (Reichle, 2012).

The National Centers for Environmental Prediction Climate Forecast System Reanalysis (CFSR; Saha et al., 2010) forms the basis of the AgCFSR climate forcing dataset, providing outputs from 1979 to present on a T382 (~38 km) horizontal grid. For AgCFSR we utilize CFSR's raw precipitation output rather than the gridded climate datasets that constrained its land-surface simulations (future versions of AgCFSR may take a more symmetrical approach similar to AgMERRA's use of MERRA-Land and CPCU). As newest generation reanalyses, both MERRA and CFSR have considerably higher spatial resolution than older reanalyses, which eliminates the need for the preliminary downscaling performed in the creation of the Princeton, WFD, and GRASP forcing datasets. The NCEP/Department of Energy Reanalysis-2 (Kanamitsu et al., 2002; which is an update to Kalnay et al., 1996, Reanalysis-1) is also included in evaluations below as an example of intermediate-generation reanalyses.

2.1.3. High-resolution precipitation products

High-resolution precipitation products (HRPP) combine information from polar-orbiting microwave instruments with geosynchronous infrared satellites to produce nearly continuous, 1/4° daily precipitation datasets (see overview and comparison with reanalyses by Ruane and Roads, 2007a). The climate forcing datasets below utilize three HRPPs in their construction: the Tropical Rainfall Measuring Mission 3B-42 product (TRMM; Huffman et al., 2007), Precipitation Estimation using Remote-Sensing and Artificial Neural Networks (PERSIANN; Hsu et al., 1997), and Climate Prediction Center Morphing Product (CMORPH; Joyce et al., 2004). TRMM, PERSIANN, and CMORPH begin in 1998, 2001, and 2003, respectively, and extend through 2010. PERSIANN and CMORPH capture precipitation equatorward of 60° N/S (covering 99.8% of major crop area) while TRMM extends poleward only to 50° N/S (sufficient to capture 91% of major crop area). GPCP's 1° daily precipitation product (v1.1; Huffman et al., 2001) from October, 1996, through August, 2009, is also utilized below in the evaluation of precipitation datasets.

2.1.4. NASA/GEWEX Solar Radiation Budget data

The Solar Radiation Budget dataset assembled by NASA and the Global Energy and Water Exchanges Project (the NASA/GEWEX SRB; Stackhouse et al., 2011) provides 1° daily incident solar radiation data globally from 1983 to 2007. These data have been distributed widely through the Agrometeorology Product of NASA's Prediction of Worldwide Energy Resources website (POWER; http://power.larc.nasa.gov). POWER provides the FlashFlux radiation data after 2007, however these data were not used as this would introduce a substantial discontinuity from 2007 to 2008. White et al. (2011b) found that SRB provides higher correlations with high-quality station measurements than do weather generation techniques over the United States, and also noted that SRB may improve upon many cooperative station measurements of solar radiation (which are often poorly maintained) or sunshine hour reports (which are inherently subjective).

2.2. Calibration and evaluation networks of meteorological stations in agricultural areas

To develop and evaluate the climate forcing datasets we independently constructed two meteorological station datasets: the first for calibration of parameters related to the diurnal temperature range and the number of rainy days (described in the next section) and then a larger second set for evaluation. These datasets were restricted to agricultural areas to emphasize the regions of expected application and to ensure that errors in high-latitude regions (where solid precipitation under-catch is often a problem; Adam and Lettenmaier, 2003) do not force compensating errors over farmed land. Agricultural areas were defined by using the 5 arcmin Monfreda et al. (2008) agricultural coverage maps, summing together land use for maize, wheat, rice, soybean, cotton, millet, sorghum, sugarcane, sugarbeet, groundnut, and barley (Fig. 1a; including additional crops does not dramatically alter agricultural areas).

2.2.1. Calibration meteorological station dataset

The calibration dataset (Fig. 1b) was generated by drawing stations from the US Historical Climate Network, the Global Historical Climate Network, the National Climatic Data Center's Global Summary of the Day, and additional stations ($\sim$7%) provided by AgMIP partners. The aim in constructing this dataset was to mimic the agricultural density with comparable station density, leading to more stations in regions with widespread agriculture, only a sampling of stations in the more sparsely farmed areas, and no stations in places like Greenland where row agriculture is not prominent. Stations were selected to have a minimal number of missing values in the 1980–2010 period, and stations not representative of their surrounding agricultural lands were removed (e.g., a high elevation station on Mount Fuji in Japan). In areas with a high density of stations we were able to locate stations with at least 90% temporal coverage for temperature and precipitation, but stations with longer gaps were included in important agricultural zones that would otherwise not be represented (none with less than 50% of daily precipitation data from 1980 to 2010). Station data quality assessment (utilizing algorithms and then hand-checks) allowed us to flag exceptionally high precipitation events, unnatural strings of consecutive values, artificially-filled data, unphysical data (e.g., rainfall < 0 or days where T_{min} exceeded T_{max}), trends and regime shifts suggesting a moving station, and temperatures that were more than 4 standard deviations from the monthly mean and not associated with physically consistent deviations in other variables and surrounding days. Nearby stations and even media reports were examined in order to corroborate high precipitation events that were not erroneous. In total, the calibration dataset includes 737 meteorological stations (49 provided by AgMIP partners).

2.2.2. Evaluation meteorological station dataset

The evaluation dataset (Fig. 1c) was drawn from the 6103 meteorological stations in the Hadley Integrated Surface Dataset (HadISD version 1.0.0.2011f; Dunn et al., 2012) according to a five step quality control process. First, stations that did not fall on the Monfreda et al. (2008) agricultural land mask (Fig. 1a) were eliminated. The HadISD dataset has undergone extensive quality control on temperatures, but no such corrections have been made to precipitation. The second step was therefore to flag years in which precipitation observations were recorded but less than 10 rainfall events occurred despite a station having more than 1000 rainfall events over the 1980–2010 period. This process eliminated stretches in which missing data were erroneously recorded as 0 mm/day measurements. As a third step, years in which less than 10 dry days were recorded were flagged as periods where observations were only taken when precipitation occurred. Fourth, rainfall events over 200 mm in a single day were eliminated, as a sub-sample revealed many of these to be spurious outliers that are potentially the result of accumulated precipitation being reported as a single day's total (for example a whole weekend of rainfall being measured on Monday). This outlier threshold removes 0.23% of total days, which undoubtedly contains several true events but is small compared to an overall 30% wet-day rate. Calculations including these high rain events resulted in overall reduced skill as would be expected when including erroneous data points, however the inclusion of these results did not affect overall patterns in skill across the considered climate datasets. Finally, each station was classified according to its temperature and precipitation coverage over the 1980–2010 period, and the top three classes were included in the evaluation dataset. The vast majority of these stations have measurements for at least 90% of the daily temperatures and precipitation, while stations with at least 80% temperature and 50% precipitation coverage were included to augment the representation of tropical regions. In total, 2324 stations are included in the evaluation dataset. While the evaluation dataset is $3\times$ larger than the calibration dataset, it is likely that several stations are present in both datasets; however calibration was restricted to universal coefficients governing the diurnal temperature range and number of rainy days (described in the next section) rather than local corrections that would give a false impression of fidelity.

It is likely that many of the stations included in the HadISD dataset were also incorporated into the construction of GPCC, CRU, and WM gridded temperature and precipitation observational datasets. Disentangling the station and gridded datasets is beyond the scope of this study, but the resulting gridded datasets contain a host of additional information (e.g., additional stations within a given grid box, interpolation rules, and weather stations in neighboring grid box) that would prevent a one-to-one match between station observations and the gridded products. The gridded observational datasets also do not contain sub-monthly information, enabling a clear comparison between the climate forcing datasets and HadISD station datasets on the daily timescale.

3. Calculation

Table 2 provides an overview of the methods utilized in the construction of each variable included in the AgMERRA and AgCFSR climate forcing datasets. Details of these procedures are provided below.

3.1. Scope and resolution of AgMERRA and AgCFSR

The AgMIP climate forcing datasets are designed to cover the 1980–2010 period, providing 30 full planting seasons even for
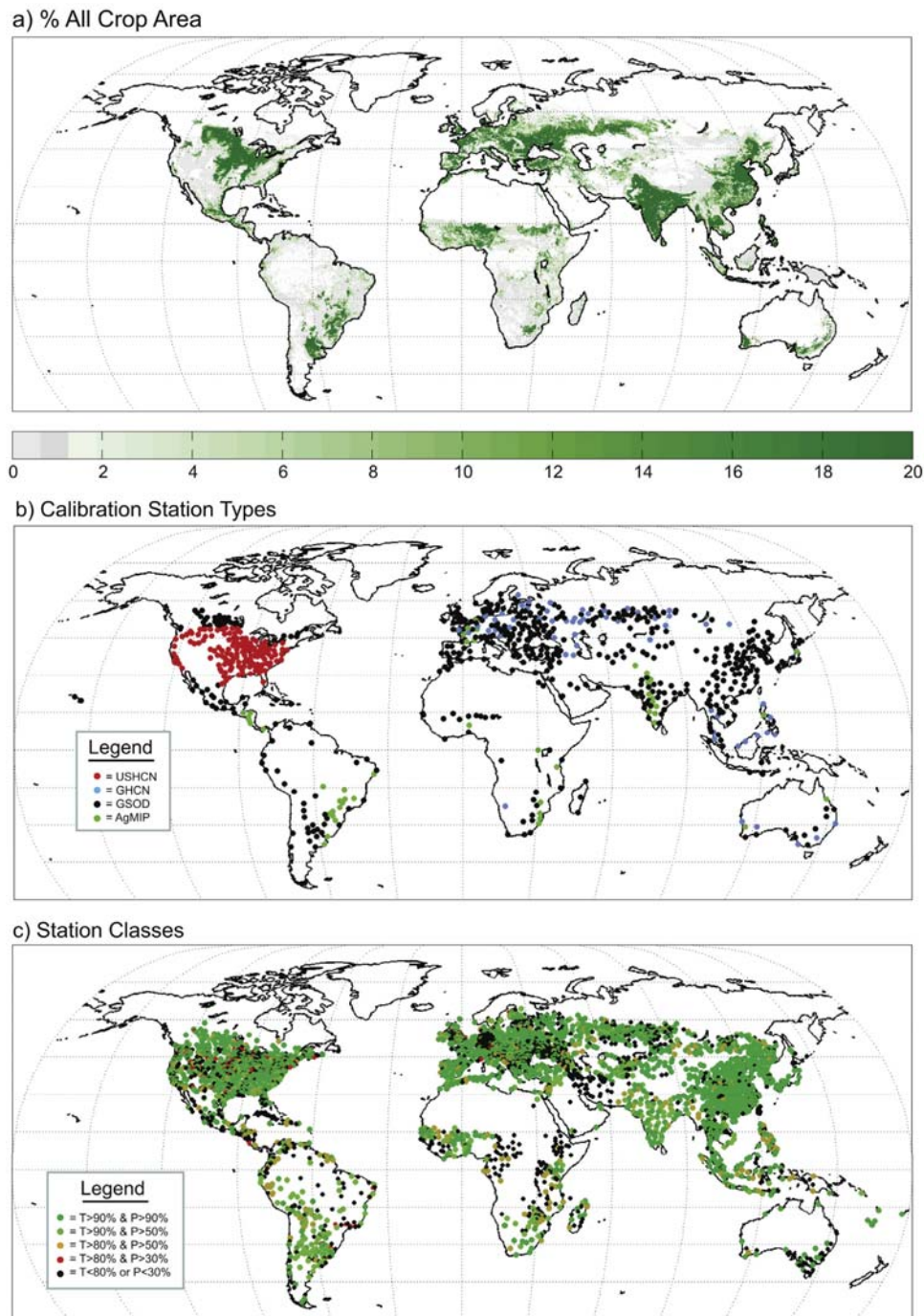
Fig. 1. Creation of a network of meteorological stations in agricultural regions for calibration and evaluation of climate forcing datasets for agricultural applications. (a) Percentage of land used for agricultural purposes (major crops from Monfreda et al., 2008; note that the darkest green color includes many areas with well more than 20% agricultural land use); (b) preliminary network of stations used for calibration (color represents source of data); (c) HadISD agricultural subset used for evaluation (color represents different levels of data coverage–red and black stations were not used in this study).

crops planted near the end of the calendar year and harvested in the next (e.g., winter wheat in the Northern Hemisphere and summer crops in the Southern Hemisphere). These 30 seasons represent the World Meteorological Organization's minimum number of years for a climatology (WMO, 1989), and thus the climate forcing datasets allow for the simulation of a full climatology of agricultural response. Data are provided at daily resolution to match the input resolution of the vast majority of crop models, and are stored on UTC rather than local time (implications of this choice are discussed

in the gap-filling section below). AgMERRA and AgCFSR contain the variables necessary for these agricultural models to function, including minimum and maximum temperature (T_{min} and T_{max}), precipitation, solar radiation, wind speed, and relative humidity (from these, secondary variables like vapor pressure or potential evapotranspiration may also be calculated).

AgMERRA and AgCFSR are stored at 1/4° horizontal resolution, although temperatures and solar radiation are derived from coarser datasets, as described below and summarized in Table 2. Data are

provided on all land areas where precipitation data were available for the construction of both AgMERRA and AgCFSR. Final coverage is constrained mostly by the CRU and MERRA-Land, excluding Antarctica and glaciated portions of Greenland but also many of the small Pacific islands. AgCFSR data do not include land areas north of 73° N and the portion of Siberia that extends into the Western Hemisphere; these areas include little agriculture. Together, the datasets include more than 99.7% of the major agricultural lands from Monfreda et al. (2008).

3.2. Maximum and minimum temperatures

AgMERRA and AgCFSR temperatures for any given day d in month m of a given year were generated in a four-step process summarized by the following equations:

$$T_{avg_{AgMERRA}}(d, m) = T_{avg_{MERRA}}(d, m) + \overline{T}_{adj}(m) \tag{1}$$

$$T_{max_{AgMERRA}}(d, m) = T_{max_{MERRA}}(d, m) + \overline{T}_{adj}(m) + \gamma \left(\text{DTR}_{CRU}(\overline{m}) - \text{DTR}_{MERRA}(\overline{m})\right) \tag{2}$$

$$T_{min_{AgMERRA}}(d, m) = T_{min_{MERRA}}(d, m) + \overline{T}_{adj}(m) - \gamma \left(\text{DTR}_{CRU}(\overline{m}) - \text{DTR}_{MERRA}(\overline{m})\right) \tag{3}$$

First, daily maximum, minimum, and average temperatures from the original MERRA and CFSR reanalyses were linearly interpolated to a 1/2° grid. Second, the average temperature for each calendar month m was calculated from each reanalysis as well as the 1/2° global gridded observational datasets provided by the Climate Research Unit (CRU TS3.1; Harris et al., 2013) and the University of Delaware (Willmott and Matsuura, 1995; WM). To reduce biases in the station density, aggregation, and interpolation of the gridded observational datasets, at each point the average of CRU and WM was calculated for each month and compared against the corresponding average monthly temperature from the interpolated reanalysis. As the WM dataset ends after 2008 and CRU was only available through 2009 at the time of calculation, the remaining years were estimated using mean biases between the gridded datasets and reanalyses in the 1980–2008 period when all were available. The difference in average temperatures for any given month, $\overline{T}_{adj}(m)$, is then added to daily average, maximum, and minimum temperatures within that month to impose the observed monthly mean.

Next, we compare the monthly average diurnal temperature range (DTR) from the reanalysis and CRU (WM DTR is not available) using days in a given calendar month from all years (denoted with an overbar for the month, $\text{DTR}_{CRU}(\overline{m})$, which is the same e.g., for July, 2004, as for July, 2009). As many agricultural models simply average the daily extreme temperatures rather than resolving the diurnal cycle, we ensure consistency by adding a fraction, γ, of the difference in DTR to the T_{max} and the same portion is subtracted from the minimum temperature. For AgCFSR γ was set to 1/2, resulting in the exact matching of CRU's DTR as was done for each of the other climate forcing datasets. Utilizing the calibration station dataset, for AgMERRA we found reduced biases in mean T_{max} and T_{min} when $\gamma = 3/8$, resulting in a final DTR that is 3/4 of the way between the DTR of MERRA and CRU (e.g., if $\text{DTR}_{CRU}(\overline{m}) = 14$ and $\text{DTR}_{MERRA}(\overline{m}) = 10$, $\text{DTR}_{AgMERRA}(\overline{m}) = 13$). The benefit of including MERRA DTR (albeit at a 1/4 weighting) suggests that MERRA's dynamical core can capture diurnal processes not captured in CRU's aggregation and interpolation procedures.

Finally, we ensure that $T_{max} > T_{min}$ on those rare days where small diurnal cycles in reanalyses are overwhelmed by differences between the mean diurnal temperature ranges of MERRA and CRU. In these cases T_{max} and T_{min} are separated by 0.4 °C about their average.

The result is a daily time series of T_{max}, T_{min}, and T_{avg} that have the reanalyses' sub-seasonal patterns and diurnal skew (whereby $[T_{max} - T_{avg}] \neq [T_{avg} - T_{min}]$ and $T_{avg} \neq [T_{max} + T_{min}]/2$ in most cases). T_{avg} has the monthly averages (and therefore interannual variability) of the global gridded observational datasets, and T_{max} and T_{min} follow with their characteristic diurnal temperature ranges. Its effective resolution comes from the 1/2° global datasets, but the 1/2° value is stored in each of four 1/4° gridboxes to match the eventual resolution of AgMERRA and AgCFSR (leading to the 1/2° effective resolution for temperatures in Table 2).

To evaluate the daily variability of T_{max} and T_{min}, the seasonal cycle was averaged across all years in each dataset and then smoothed with a 15-day averaging filter. After removing this seasonal cycle we are able to compare daily anomalies between the evaluation dataset, the climate forcing datasets, and the reanalyses, in addition to comparisons of mean bias.

3.3. Precipitation

The AgMIP climate forcing datasets are designed to take advantage of the reanalyses' recognition of large-scale conditions susceptible to precipitation events while recognizing that reanalysis parameterizations struggle to capture rainfall frequencies, distributions, and totals (Bosilovich et al., 2008; Lorenz and Kunstmann, 2012). AgCFSR, like each of the existing climate forcing datasets, begins with reanalysis precipitation (from CFSR) that does not include any precipitation assimilation. AgMERRA, however, utilizes the MERRA-Land precipitation dataset that incorporates precipitation observations from the CPC (Reichle, 2012).

AgMERRA and AgCFSR precipitation adjust the original reanalysis time series in a four step process. First, the original daily time series is linearly interpolated to the 1/2° CRU grid (MERRA-Land is missing for ocean points, so some coastal regions were re-gridded using nearest neighbor interpolation).

The second step adjusts daily precipitation events (defined as those with at least 0.1 mm). For AgCFSR these are shifted to match the number of precipitation days in that particular month indicated by the CRU TS3.10 dataset (Harris et al., 2013; missing 2010 wet days estimated from 1980 to 2009 overlap when reanalyses and CRU were available). For AgMERRA the calibration dataset indicated that the best result occurs when the number of rainy days was set to the average of the wet days in CRU and those in MERRA (defined as those with at least 0.5 mm; using a 0 mm wet/dry threshold for reanalyses results in too many rainy days). If the re-gridded reanalysis had too many precipitation days, amounts for the equivalent number of days with the lowest precipitation totals were changed to zero. If additional precipitation days were required, 0.3 mm rainfall events were added for the necessary number of days beginning with those with the least solar radiation (indicating the presence of clouds on a day where precipitation was not simulated). The GRASP dataset was generated with a similar procedure for adding and removing rainy days, while the WATCH datasets adjusted the number of precipitation days downward but did not create any additional precipitation days to overcome monthly shortfalls.

Monthly precipitation totals from the re-gridded and wet-day-corrected reanalysis data are then compared against the ensemble average of three 1/2° gridded observational products (CRU, WM, and the Global Precipitation Climatology Centre Full Data Product version 6, GPCC, Schneider et al., 2011) to produce an adjustment factor multiplied by each day in that month. This results in an adjusted value at 1/2° resolution ($P'_{AgMERRA}(d, m)$), as described for

AgMERRA by Eq. (4) (AgCFSR uses CFSR in a similar manner):

$$P'_{AgMERRA}(d, m) = P_{MERRAland}(d, m)$$

$$\times \frac{1}{3}\left(\frac{\left(\overline{P}_{CRU}(m) + \overline{P}_{WM}(m) + \overline{P}_{GPCC}(m)\right)}{\overline{P}_{MERRAland}(m)}\right). \quad (4)$$

Although GPCC data were available over the entire 1980–2010 period, the WM dataset ends after 2008 and CRU was only available through 2009 at the time of calculation. 2009 and 2010 were therefore estimated using mean biases between the gridded datasets in the 1980–2008 period when all were available. An ensemble of multiple gridded observational datasets was utilized in order to take advantage of offsetting biases in their aggregation and interpolation algorithms. Although it is likely that particular products perform best in specific regions, no product is clearly superior in all regions and global consistency is preferable to a patchwork of datasets.

The final step utilizes the suite of high-resolution precipitation products (HRPP) to achieve additional resolution in precipitation. Although shorter than a full 30-year climatology, these HRPP provide enough years that their mean differences for any given calendar month capture fine-scale differences due to land cover, coastlines, and terrain without being overwhelmed by particular storm events. For each $1/2°$ gridbox from the gridded observational products, the ensemble of TRMM, CMORPH, and PERSIANN provide four $1/4°$ gridboxes ($i = 1,2; j = 1,2$). Precipitation scaling factors are then calculated according to each grid box's fraction of the $1/2°$ aggregated value. These scaling factors are then applied to all days within a given calendar month (e.g., all March days are multiplied by the same scaling factors regardless of year), adding geographical detail at $1/4°$ resolution as described by the following equation:

$$P_{AgMERRA}(d, m, i, j) = P'_{AgMERRA}(d, m)$$

$$\times \frac{\left(\overline{P}_{TRMM}((, i, j) + \overline{P}_{CMORPH}(\overline{m}, i, j) + \overline{P}_{PERSIANN}(\overline{m}, i, j)\right)}{\sum_{i,j}\left(\overline{P}_{TRMM}(\overline{m}, i, j) + \overline{P}_{CMORPH}(\overline{m}, i, j) + \overline{P}_{PERSIANN}((\overline{m}, i, j)\right)}. \quad (5)$$

The above procedure describes the final creation of AgMERRA and AgCFSR precipitation, however many other methodologies were evaluated against the calibration station dataset and ultimately found to fall short of desired results. One noteworthy approach employed a quantile-mapping approach to adjust the mean of the distribution of precipitation events while holding constant the shape parameter of its fitted gamma distribution (Wilks, 1995). A second approach used quantile-mapping to adjust the reanalysis precipitation to match that of a gamma distribution fit to the HRPP datasets. These approaches ultimately failed because it was too difficult to maintain the overall integrity of the multi-year gamma distribution while also forcing specific monthly totals to match the gridded observations, and precipitation at many locations was better described by a form other than the gamma distribution. The second approach had the added challenge of overcoming substantial and fundamental differences in the shape parameters of the fitted gamma distributions for the reanalyses and HRPP.

In addition to the mean biases, we evaluate sub-seasonal variability of precipitation using statistical methods reflective of the probability of occurrence for various events (Wilks, 1995). For wet days (precipitation > 1 mm) we employ the hit rate (HR_1) defined by the following equation:

$$HR_1 = \frac{DD + WW}{DD + WW + DW + WD} \times 100\% \quad (6)$$

where DD represents the number of days that were dry in both the climate product and observations (across all evaluation dataset locations), WW the number of days that were wet in both the climate dataset and observations, and the remaining days are either wet in the climate dataset and dry in observations (WD; sometimes referred to as *false alarms*) or vice-versa (DW). Hit rate may therefore be understood as the percentage of correct wet or dry representations out of the total number of days. For more extreme precipitation events (precipitation > Q mm) we account for the fact that a persistent dry forecast would give a false appearance of skill in the hit rate, and instead utilize the threat score (TS_Q) defined by the following equation:

$$TS_Q = \frac{WW}{WW + DW + WD} \times 100\%, \quad (7)$$

where each of the events are tested against a threshold of Q mm. The threat score may therefore be understood as the percentage of days where the climate dataset correctly identifies a precipitation event compared to the total number of days where the precipitation event is either anticipated by the climate dataset and/or actually observed.

3.4. Solar radiation

Crop models require accurate solar radiation to drive their simulation of photosynthesis and the carbon balances that govern plant growth. Although the CFSR and MERRA reanalyses contain an equivalent downward shortwave radiation flux, the variable is not assimilated and is subject to biases from cloud parameterizations that remain among the largest challenges in numerical weather prediction. Following White et al. (2008), AgMERRA and AgCFSR utilize the NASA/GEWEX SRB solar radiation whenever it is available (July, 1983 through 2007). Rather than use only the monthly mean SRB values to adjust daily solar radiation time series as was done in the other climate forcing datasets, AgMERRA and AgCFSR directly utilize the SRB data after linear interpolation to a $1/4°$ grid.

To fill in the periods when SRB data are not available (1980-June, 1983 and 2009–2010), downward shortwave radiation flux from the original reanalysis was first linearly interpolated to a $1/2°$ grid. As shortwave radiation cannot be negative and is capped by astronomical limitations (determined by latitude and Julian day), we fit a beta distribution (Wilks, 1995) to the SRB and re-gridded reanalysis for each month (e.g., one SRB distribution describing 806 July days from 1983 to 2007). Using maximum solar radiation to scale the Beta distribution described by p and q parameters at each location, solar radiation from the reanalysis was shifted (using quantile-mapping) to match the properties of the SRB distribution in years when SRB data were not available. In some high-latitude locations the p parameter was capped at 200 to offset poor distributions in months when the sun set for the winter or re-emerged in the spring, with only small errors due to the low maximum radiation in these months.

3.5. Relative humidity at T_{max} and 2-m wind speed

Although required for only a substantial subset of crop models, a measurement of near-surface atmospheric moisture and wind speeds allow most models to utilize more advanced evapotranspiration (ET) parameterizations that estimate turbulent moisture fluxes in the crop environment. These variables also have applications related to the emergence and spread of agricultural pests and diseases. For each of these purposes the biophysical response is dependent most directly on vapor pressure deficit (VPD; the difference between saturated VP and actual VP), however VPD is rarely measured directly. Relative humidity (the ratio of actual VP/saturated VP) serves as a suitable proxy but experiences a large diurnal cycle as temperature variation causes large swings in the saturated VP that often overwhelm ET contributions to actual VP.

Relative humidity typically peaks near sunrise (when temperatures are at their coldest) and then falls to a minimum in the hottest part of the day before ET and cooling temperatures reverse the decline (Ruane and Roads, 2007b). As the crop models require daily inputs, AgMIP discussions encouraged the creation of a dataset for "relative humidity at maximum temperature" ($RH_{T_{max}}$; recorded at 2 m elevation), which can be converted to vapor pressure and dewpoint temperature because they correspond with a specific temperature. $RH_{T_{max}}$ approximately corresponds to the time of peak ET (and minimum relative humidity), and also mimics the ~2 pm local time sling psychrometer observations that were often used to estimate VPD when the crop models were being developed.

AgMERRA and AgCFSR rely on their original reanalyses for $RH_{T_{max}}$ and wind speed as moisture and wind observations in the free atmosphere are included in the assimilation procedures of MERRA and CFSR. The assimilated observations tend to come from above the crop canopy level, however, resulting in potential biases and differences in temporal variation between reanalyses at the 2-m level (Ruane and Roads, 2007b). As neither reanalysis directly records relative humidity in their output, we calculated $RH_{T_{max}}$ using the specific humidity and surface pressure corresponding to T_{max} (Curry and Webster, 1999) and then linearly interpolated $RH_{T_{max}}$ to the final 1/4° grid. Wind speeds from MERRA and CFSR were likewise interpolated to the 1/4° grid for AgMERRA and AgCFSR, the latter following a reduction of wind speed by 25.2% to estimate 2 m wind speed from the 10 m value in agricultural conditions (Allen et al., 1998).

4. Results

4.1. Maximum and minimum temperatures

Figs. 2 and 3 present key diagnostics for AgMERRA and AgCFSR T_{max} and T_{min} validated against the HadISD-based dataset (Figs. 2a and 3a) and compared against other climate forcing datasets and reanalyses. AgMERRA and AgCFSR have nearly identical monthly average temperatures (differing only for 2010 when neither CRU nor WM were available), however the diurnal temperature range adjustments (γ in Eqs. (2) and (3)) lead to slight differences in T_{max} and T_{min}. Comparisons between AgMERRA/MERRA and AgCFSR/CFSR reveal the improvements gained through the methodologies above.

The spatial pattern of biases (Fig. 2c and d) shows a general warm bias in AgMERRA and AgCFSR T_{max}, with the largest biases in countries where station density is low and in mountainous areas where complex topography is not resolved at coarser grid scales. The histogram of mean T_{max} biases (Fig. 2b) offers another visualization of the overall warm bias, with AgMERRA and AgCFSR peaking at +0.5 °C along with the Princeton and WATCH datasets. The warm bias is therefore likely due to a bias between the HadISD data and the gridded observational datasets (CRU and WM) used in the construction of the climate forcing datasets. AgMERRA's γ reduces the warm bias slightly in comparison to AgCFSR, Princeton, and the WATCH datasets, and each of these has a substantially tighter distribution with reduced extreme biases compared to the coarser GRASP climate forcing dataset and the reanalyses.

Resolution of sub-seasonal T_{max} in AgMERRA and AgCFSR is largely dependent on the underlying reanalyses (MERRA and CFSR), as adjustments from CRU and WM constrain only the monthly timescale. Fig. 2e shows a histogram of Pearson's correlations (r) between each climate dataset and the 2324 stations in the evaluation dataset on a daily timescale after the average seasonal cycle has been removed. AgMERRA, AgCFSR, MERRA, CFSR, and the WATCH datasets all group together tightly with very

high correlations (peaking near $r = 0.9$), with AgCFSR correlations slightly ahead of WFD among the highest two. The products based on coarser reanalyses have lower correlations, with GRASP and the R2 forming a second, slightly wider group peaking near $r = 0.85$ and the Princeton correlations peaking at $r = 0.6$. Daily correlations are highest in the mid- and high-latitudes (Fig. 2f), likely due to the reanalyses' relative comfort with synoptic patterns as opposed to tropical climates. AgCFSR improves slightly upon AgMERRA daily correlations in much of the world (Fig. 2g), although AgMERRA correlations are higher in many of the tropical areas (where correlations in both products tend to be lower). AgMERRA, AgCFSR, and the WATCH datasets have the lowest root-mean-squared difference against the evaluation dataset (RMSD near 2.6 °C), suggesting superior performance with regards to the combination of mean biases and correspondence in sub-seasonal variability (Fig. 2h).

AgMERRA and AgCFSR generally have slightly negative T_{min} biases (Fig. 3c and d). As was noted for T_{max}, larger biases occur in regions where meteorological stations are less dense and where mountain and valley stations are not adequately represented by large grid boxes. The tight distributions of AgMERRA, AgCFSR, Princeton, WFD, and WFD-EI biases all peak at −0.5 °C, again suggesting that there is a noteworthy difference between the gridded observational datasets and the evaluation dataset of Had-ISD stations (Fig. 3b). The AgMERRA distribution is again closest to zero bias, benefiting from the combination of MERRA and CRU diurnal cycles.

Histograms of T_{min} daily correlations (Fig. 3e) look very similar to those of T_{max}, however the WFD-EI and especially the WFD-EI have substantially more stations in the $r = 0.95$ bin. AgMERRA and AgCFSR both peak at $r = 0.9$, with the vast majority of stations having $r > 0.7$. Once again daily correlations are highest in the mid- and high-latitudes (Fig. 3f). Patterns of differences between AgCFSR and AgMERRA's correlations are also accentuated (Fig. 3g), with AgCFSR better outside of the tropics (where correlations in both products tend to be highest) and AgMERRA higher in the tropics (where correlations tend to be lowest). The improvement from MERRA to AgMERRA (and from CFSR to AgCFSR) is clear in the T_{min} RMSD (Fig. 3h), which again places AgMERRA and AgCFSR with WFD and WFD-EI as the top-performing climate datasets (mean RMSD near 2.6 °C).

The warm bias in T_{max} and cool bias in T_{min} combine to overestimate the diurnal temperature range for all climate forcing datasets, while the original reanalyses are in closer balance with the HadISD DTR but have a much wider spread (Fig. 4a). This suggests that the γ factor in Eqs. (2) and (3) would have been higher had the HadISD dataset been used in the calibration process, resulting in overall reductions in DTR. The T_{max} and T_{min} biases compensate in a convenient manner for AgMERRA and AgCFSR, peaking tightly at 0 °C average above the Princeton, WFD-EI, and WFD datasets (Fig. 4b). Although these climate forcing datasets' DTRs are too high, the compensating biases suggest that the bias for daily average temperatures would be lower than either the T_{max} or T_{min} biases and that the products capture an appropriate diurnal cycle as represented by the ratio of $(T_{avg} - T_{min})/DTR$.

4.2. Precipitation

Precipitation diagnostics for AgMERRA and AgCFSR precipitation are compared with the evaluation dataset and other climate products in Fig. 5. Due to the combination of unrealistically frequent extreme events and numerous missing days reported in many of the HadISD stations' precipitation records, the mean precipitation rate was determined to be too problematic as a basis for climate product validation. To illustrate this, Fig. 5a shows the mean annual precipitation from AgMERRA and Fig. 5c presents its biases against the evaluation dataset. While these biases are low over much of
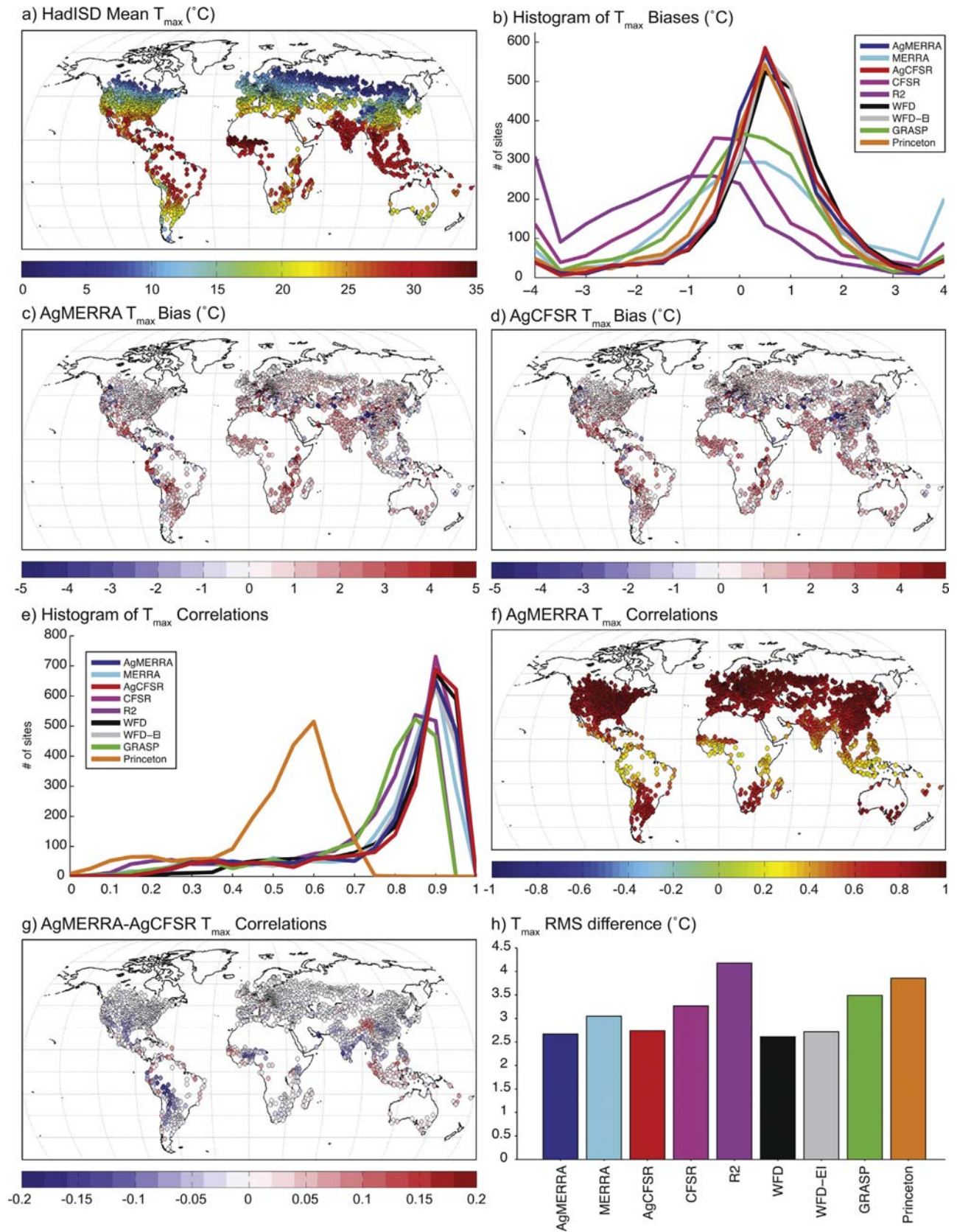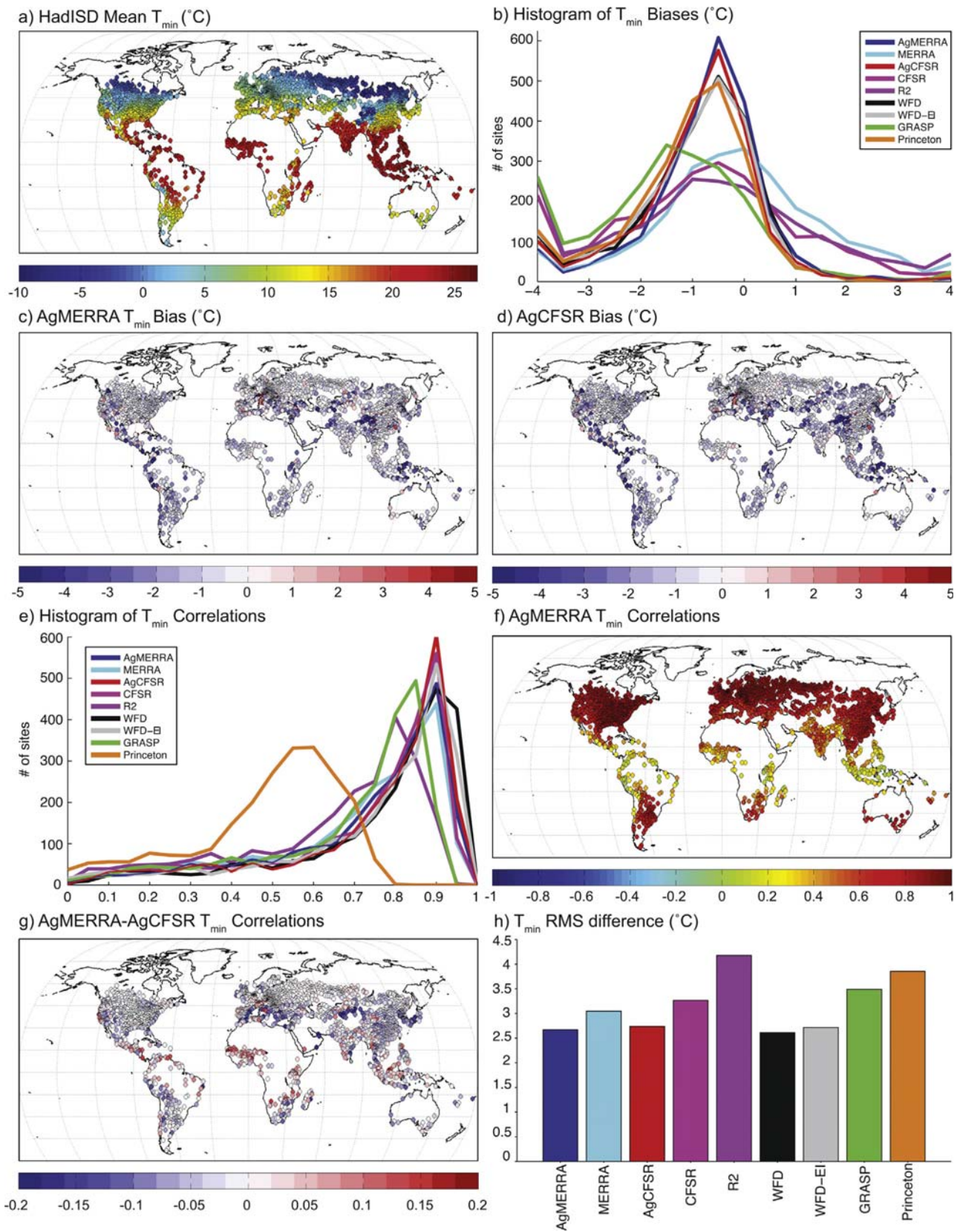
Fig. 2. Diagnostics of AgMERRA and AgCFSR T_{max} against 2324 stations in HadISD-based evaluation dataset (2324 stations) and against other climate forcing datasets and reanalyses. (a) Mean T_{max} from HadISD-based evaluation dataset; (b) Histogram of mean T_{max} bias; (c) AgMERRA mean T_{max} bias; (d) AgCFSR mean T_{max} bias; (e) Histogram of Pearson correlations for daily T_{max} (with seasonal cycle removed); (f) Geographical pattern of AgMERA T_{max} correlations; (g) AgMERRA–AgCFSR differences in T_{max} correlations; (h) Root-mean-squared difference from daily T_{min} series. Note that the left-most and right-most bins in the histogram contain all values beyond the limits of the x-axis.

Fig. 3. Diagnostics of AgMERRA and AgCFSR T_{min} against 2324 stations in HadISD-based evaluation dataset (2324 stations) and against other climate forcing datasets and reanalyses. (a) Mean T_{min} from HadISD-based evaluation dataset; (b) Histogram of mean T_{min} bias; (c) AgMERRA mean T_{min} bias; (d) AgCFSR mean T_{min} bias; (e) Histogram of Pearson correlations for daily T_{min} (with seasonal cycle removed); (f) Geographical pattern of AgMERRA T_{min} correlations; (g) AgMERRA-AgCFSR differences in T_{min} correlations; (h) Root-mean-squared difference from daily T_{min} series. Note that the left-most and right-most bins in the histograms contain all values beyond the limits of the x-axis.
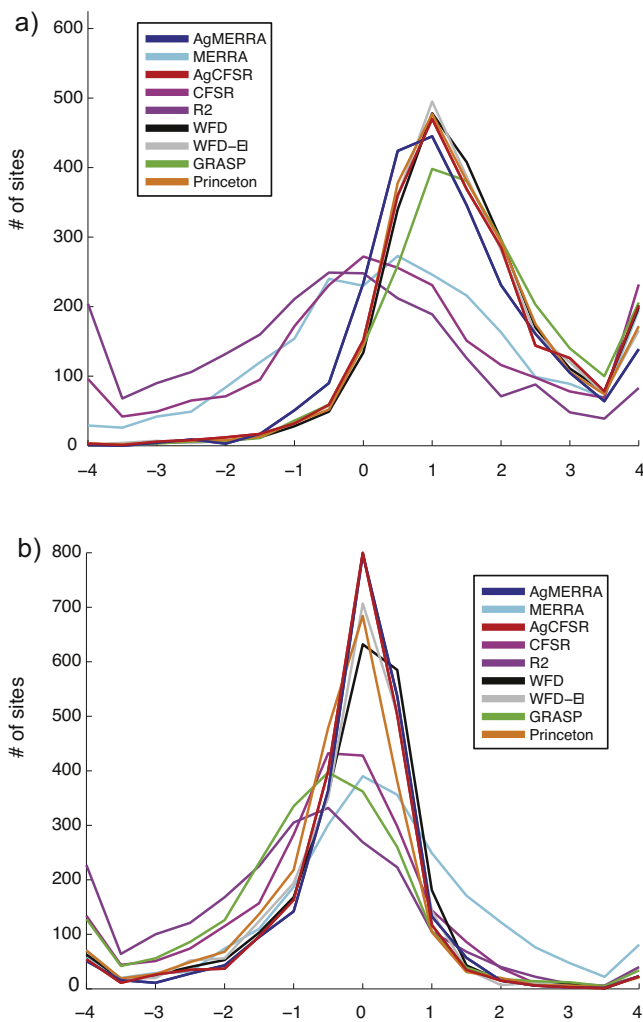
Fig. 4. Diurnal temperature range comparisons. (a) Mean DTR bias; (b) Average of T_{max} and T_{min} biases. Note that the left-most and right-most bins in the histogram contain all values beyond the limits of the x-axis.

North America where station precipitation quality control has been most extensive, in other parts of the world there is a substantial dry bias.

Fig. 5b shows a histogram of percentage differences between AgMERRA and the other climate forcing datasets at each of the evaluation station locations, and includes the gridded observational datasets (CRU, WM, and GPCC) to compare against established observational datasets while also recognizing observational uncertainty. AgCFSR matches AgMERRA due to the identical imposition of monthly precipitation totals from the ensemble of gridded climate products, and therefore matches the climatology presented in Fig. 5a. The gridded observational datasets tightly cluster around their ensemble average, with differences within 10% for the vast majority of sites examined. MERRA-Land and the WATCH forcing datasets have a slightly wider distribution of mean differences, with MERRA-Land peaking at zero difference and the WATCH datasets peaking with slightly wetter conditions. This difference is approximately the size of the adjustments to the CRU precipitation made by WATCH (but not AgMERRA or AgCFSR) to account for the "under-catch" of solid precipitation (Adam and Lettenmaier, 2003), however this correction was also made for the Princeton dataset, which is only slightly wetter than AgMERRA and AgCFSR and peaks tightly at zero difference. GRASP and the coarser

reanalyses have a very wide distribution of mean differences, indicating both wetter and drier regions are common. Note that here we evaluate percentage differences rather than precipitation totals, which would appear small in arid regions even with high percentage differences. This more sensitive metric also identifies that the coarser reanalyses and GRASP have a large number of sites where precipitation is at least 50% higher than is captured in the construction of AgMERRA.

Although the HadISD dataset proved unsuitable for mean precipitation validation, the daily observations are helpful in determining the sub-seasonal character of precipitation across the evaluation dataset. These variations are largely determined by the underlying reanalyses, although adjustments in the number of rainy days and monthly totals also affect the frequency and intensity of precipitation events. Fig. 5d presents a histogram of the correlation of daily precipitation in each climate product compared to the HadISD station data. AgMERRA's correlations are substantially higher than any of the other climate forcing datasets, peaking in the $r = 0.8$ bin with a substantial number of stations in the $r = 0.85$ and $r = 0.9$ bins and very few stations having correlations below $r = 0.2$. These results are particularly encouraging given the large spatial variability in precipitation and the likelihood that some precipitation events are not well captured by a sparse network of stations (Dzotsi et al., 2013). AgMERRA's performance is clearly the result of its basis in MERRA-Land, which has the best overall performance through a combination of MERRA's simulation of the water cycle and MERRA-Land's additional incorporation of CPC precipitation data. AgCFSR, WFD, WFD-EI, GRASP, MERRA, and CFSR all peak at $r = 0.7$ with only a small number of stations reaching $r = 0.85$. The older generation R2 and the coarse satellite product GPCP peak at $r = 0.4$. The Princeton dataset is omitted from these sub-monthly precipitation metrics because it utilizes a resampling approach that was not designed to capture specific daily precipitation events within a given month. Including the 200+ mm rain events that were eliminated as untrustworthy observations (or using higher thresholds) reduces each of these correlations but does not affect the overall pattern of AgMERRA having highest correlations and AgCFSR landing with the other climate forcing datasets. AgMERRA daily precipitation correlations tend to be highest in areas with the densest station coverage, suggesting that data quality in the observational dataset is also a potential limitation on reaching high correlations (Fig. 5e). AgMERRA has higher correlations than AgCFSR in nearly all regions with the prominent exception of Argentina, where correlations are comparable to AgCFSR (Fig. 5f).

AgMERRA also follows MERRA-Land as the top performing climate products for the wet day hit rate (Fig. 5g), correctly identifying whether a day was wet 83.0% and 83.6% of the time, respectively. WFD (80.5%), GRASP (79.8%), AgCFSR (79.8%), and WFD-EI (78.6%) form the next group of high-performing products, with the other reanalyses and the GPCP below 77%. Due to the stochastic nature of sub-monthly precipitation in the Princeton data, its hit rate is slightly below the 70% mark that would be achieved by assuming that each day was a dry day.

AgMERRA also follows MERRA-Land to achieve the top performance among climate forcing datasets based on threat scores for all (>1 mm), at least moderate (>25 mm), and heavy (>50 mm) precipitation events (Fig. 5h). This relative performance also increases as events become more intense, with AgMERRA's threat score (53.9%) approximately 10% higher than the threat score of the next best climate forcing dataset (WFD; 48.3%) for events greater than 1 mm, 43% higher for events greater than 25 mm (22.8% compared to AgCFSR's 15.9%), and 45% higher for events greater than 50 mm (10.9% compared to AgCFSR's 7.5%). Threat scores were not significantly affected by the elimination of 200+ mm events, with differences on the order of 0.01%.
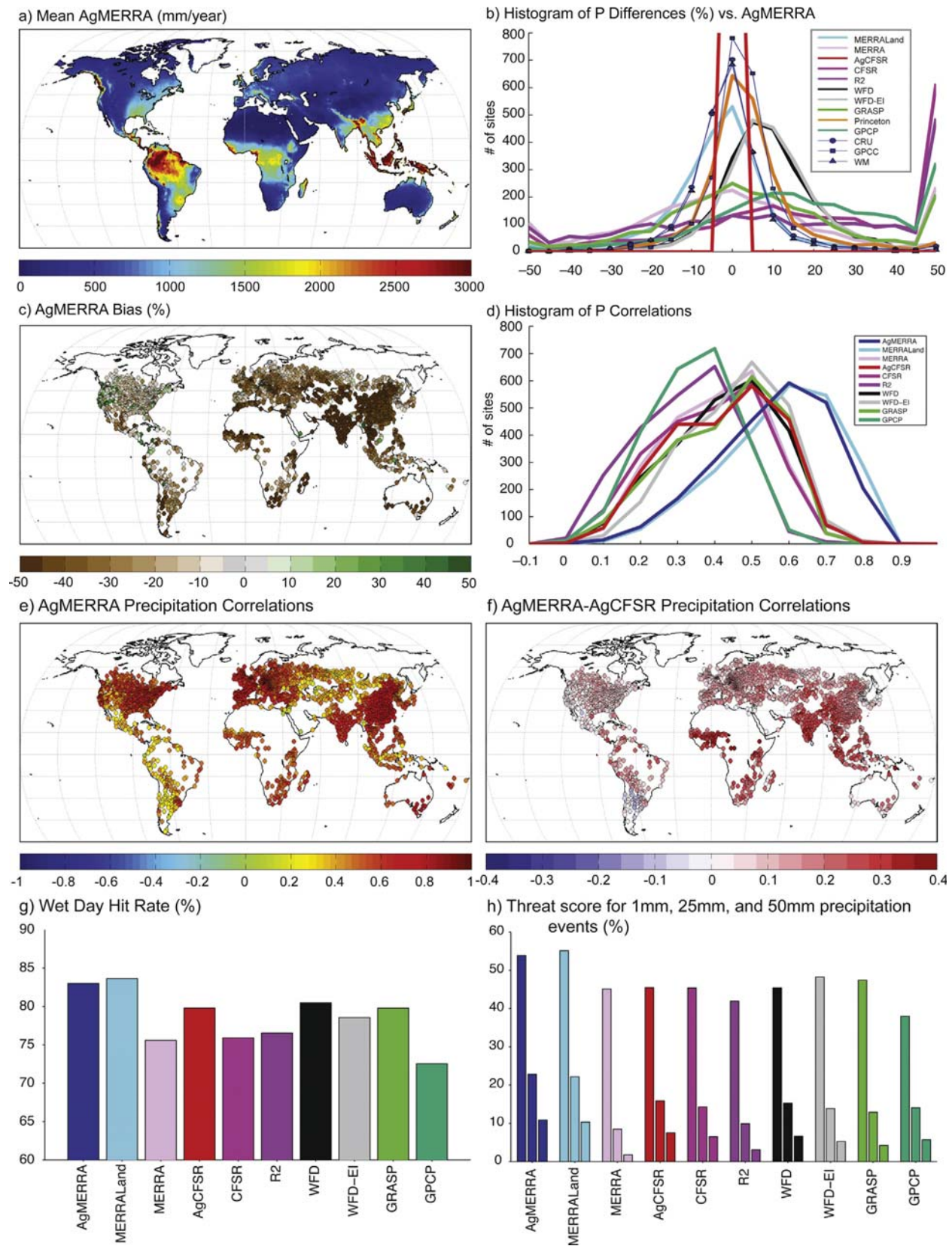
Fig. 5. Diagnostics of AgMERRA and AgCFSR precipitation against 2324 stations in HadISD-based evaluation dataset and against other climate forcing datasets and reanalyses. (a) Mean AgMERRA precipitation; (b) Histogram of mean precipitation differences against AgMERRA; (c) AgMERRA mean precipitation bias against evaluation dataset; (d) Histogram of Pearson correlations for daily precipitation against evaluation dataset; (e) Geographical pattern of AgMERRA precipitation correlations; (f) AgMERRA-AgCFSR differences in daily precipitation correlations; (g) Hit rate of precipitation days in evaluation dataset; (h) Threat scores for each dataset's 1 mm (left), 2 mm (center), and 50 mm (right) daily precipitation events in evaluation dataset. Note that the left-most and right-most bins in the histograms contain all values beyond the limits of the x-axis, and that the Princeton dataset was omitted from panels (d,g,h) because the current version resamples sub-monthly precipitation.

4.3. Solar radiation

Fig. 6a presents the mean climatology of solar radiation from AgMERRA, which is mostly a reflection of the NASA/GEWEX SRB observational product that makes up the bulk of its values and constrain the other years. AgCFSR is nearly identical, differing only at the beginning and end of the 1980–2010 period when SRB data were not available. Similarities between these two climate products are evident in the histogram of differences (vs. AgMERRA) presented in Fig. 6b, which shows AgCFSR mean differences less than 0.25 MJ/m2/day at almost every location. The Princeton and GRASP datasets are also very similar to the AgMERRA data, with less than a hundred Princeton sites showing differences near 0.5 MJ/m2/day in either direction and the GRASP data also tightly distributed around zero difference. WFD, WFD-EI, and the reanalyses have a much wider distribution, indicating substantial differences in comparison to the SRB data. WFD generally shows a negative (cloudier) difference, while WFD-EI and the reanalyses have a positive (brighter) difference at most stations (Bosilovich et al., 2011).
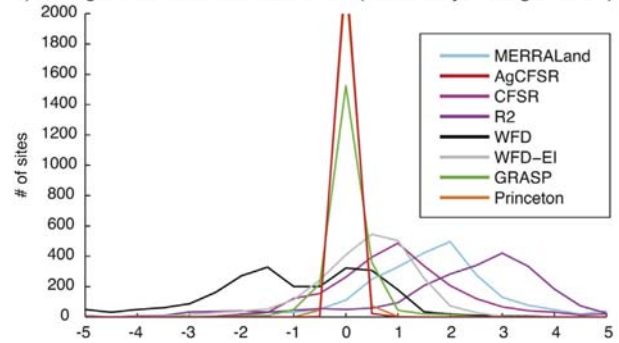
4.4. Relative humidity at T_{max} and wind speed

$RH_{T_{max}}$ and wind speed are taken nearly directly from the MERRA and CFSR reanalyses. The exact mechanisms for differences in CFSR and MERRA are beyond the scope of this study (Bosilovich et al., 2011; Meng et al., 2012, provide more detail on MERRA and CFSR, respectively), but differences are a result of the simulation of boundary-layer profiles as both products assimilate nearly the same observations of the free atmosphere. The mean climatologies are presented here as references for future work. Fig. 6c and d show very similar patterns in mean $RH_{T_{max}}$ for AgMERRA and AgCFSR, respectively. AgMERRA tends to have higher $RH_{T_{max}}$ over the tropics, most notably in Mesoamerica and the Amazon, West Africa, and Indonesia. Wind speeds (Figs. 6e and f) demonstrate larger differences, although the most dramatic differences are in areas with little agricultural production (AgMERRA wind speeds are greater at high latitudes and over major deserts). Differences of greater interest to the agricultural modeling community include lower wind speeds over Eastern North America in AgMERRA, higher winds over Northern Europe in AgMERRA, and lower
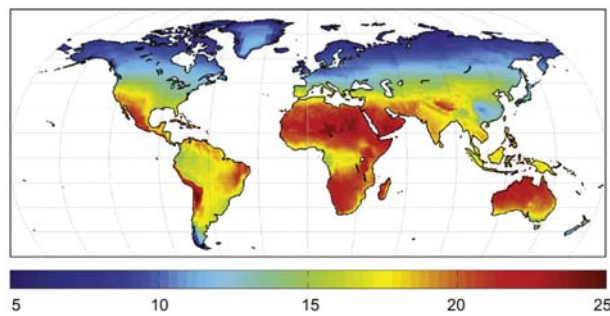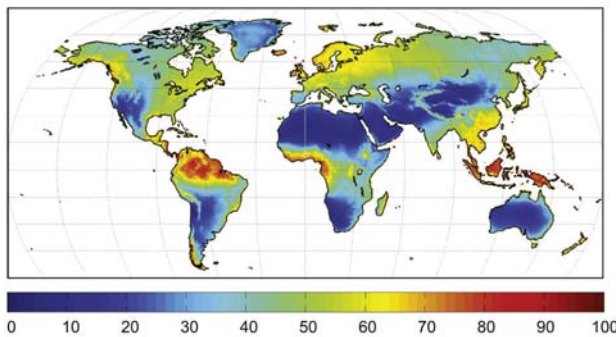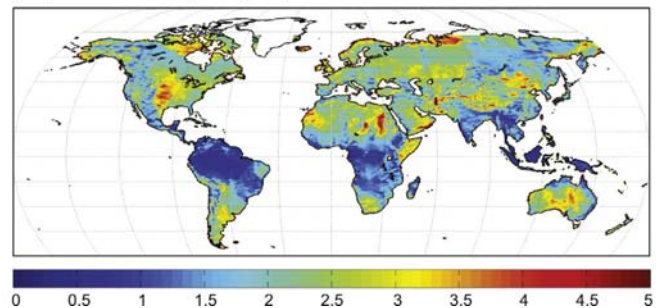

Fig. 6. Overview of AgMERRA and AgCFSR climatologies for radiation, relative humidity at T_{max}, and wind speeds. (a) Mean AgMERRA solar radiation; (b) histogram of solar radiation biases in comparison to AgMERRA; mean $RH_{T_{max}}$ for (c) AgMERRA and (d) AgCFSR; mean wind speed for (e) AgMERRA and (f) AgCFSR. Note that the left-most and right-most bins in the histogram contain all values beyond the x-axis.

surface wind speeds associated with the Asian Monsoon regions in AgCFSR.

5. Discussion

5.1. AgMERRA and AgCFSR advantages

Evaluating across all metrics, both AgMERRA and AgCFSR emerge as strong, novel climate forcing datasets that are appealing for application and further development. The AgMERRA dataset, however, has substantial advantages in its daily precipitation performance that recommend it most highly for immediate use.

Much of the procedure for AgMERRA and AgCFSR is drawn from the work that developed the Princeton, WFD, WFD-EI, and GRASP datasets, however several distinguishing features promote their use for agricultural applications. As was done for other climate forcing datasets, AgMERRA and AgCFSR utilize gridded observational datasets to remove biases that are important to agricultural production in the interannual variation and mean seasonal cycle of temperature, precipitation, and solar radiation. AgMERRA and AgCFSR use an ensemble of the gridded observational datasets for temperature and precipitation, however, acknowledging the uncertainties related to these datasets in sparsely-observed regions, and draw from an ensemble of high-resolution precipitation products to capture enhanced spatial resolution. AgCFSR is unique in its use of the most recent NCEP-based reanalysis system, with higher original resolution and improved dynamics over earlier generation reanalyses. AgMERRA also uses a modern generation of reanalysis that has not previously been developed into a climate forcing dataset, featuring MERRA-Land daily precipitation variation that demonstrates substantially higher correspondence with observations than is seen in other climate forcing datasets on this crucial agro-climatological variable.

The selection of a climate dataset for any particular application requires a process to match the agro-climatic properties of greatest importance to the investigation and any model used. AgMERRA's sub-monthly precipitation fields make it most appealing for simulations of rain-fed agriculture and the implications of water stress, extreme events, and changing precipitation patterns. AgCFSR has slightly better sub-monthly temperature fields, which may be of most interest for studies related to heat stress or irrigated conditions. Both datasets also offer high-quality SRB radiation and are unique in providing humidity data synchronized with the maximum temperature time of day to better resolve the diurnal cycle of near-surface moisture for evapotranspiration and water stress studies. Of course, the benefits of these improvements will only be reflected by agricultural models that are accurately sensitive to these features. Crop and livestock model responses to extreme events continue to be a major focus of AgMIP, and it is likely that continuing model improvement will further highlight differences in the climate forcing datasets.

5.2. AgMERRA and AgCFSR limitations

AgMERRA and AgCFSR combine data from reanalyses with global observational datasets to form a best estimate of the daily climate over the 1980–2010 period. The climate forcing datasets are therefore expected to be most accurate for regions and atmospheric processes where the underlying datasets are least biased. Errors in the reanalyses' simulation of complex dynamics (particularly around convection and moisture fluxes throughout the crop and boundary layers), the resolution of sub-grid-scale features (particularly in mountainous regions), and interpolation and assimilation of a sparse network of meteorological observations (particularly, but not exclusively, in developing countries) likely

manifest themselves in the AgMIP climate forcing datasets. For example, synoptic weather patterns over the dense observational networks of the United States and Europe are likely to be better captured than convective rain events over mountainous portions of Eastern Africa. These limitations are common to each of the climate forcing datasets compared here.

AgMERRA and AgCFSR depend on datasets that are not static through time, so care must be taken in analyzing long-term trends. The gridded climate datasets that provide monthly values may be affected by a changing number of nearby meteorological stations in any given region, which may alter the gridded value and nearby interpolated values. Various satellite instruments also launched over the 1980–2010 period, altering the types and quality of remote observations assimilated into MERRA and CFSR. These changes may affect long-term trends in relative humidity and wind speed data, and may introduce subtle changes in the sub-monthly pattern of temperature and precipitation events. For all of the above reasons it is important that AgMERRA and AgCFSR be considered as climate information records, not climate observations. While these datasets were created to allow the simulation of agricultural production and trends over the 1980–2010 period, strictly climatological trend analysis will reflect the underlying observational datasets rather than unique contributions of these blended products.

The combination of datasets and largely independent adjustment methodologies can lead to unphysical variable relationships that may be problematic for certain applications. For example, separate adjustments to temperature, rainfall, and relative humidity combine to throw off the balance of water and energy in the original data. Caution must therefore be used in any application of AgMERRA or AgCFSR data that solves for missing water and energy budget components by assuming a closed water or energy budget. The correlation between precipitation days and solar radiation (which is negative in observations and strongest in the GRASP dataset) is also degraded slightly as reanalysis radiation data are replaced with SRB data (see Reichle et al., 2011, for additional information about MERRA's rainfall-sunlight relationship). Higher negative correlations in the reanalyses follow a poorer correspondence with observed precipitation and solar radiation, as shown in the previous section.

5.3. Gap-filling applications

A common challenge for agricultural modelers is the need to fill in data gaps in meteorological station records to allow continuous simulations for a given region when simple interpolation would not be sufficient (gaps > 4 days). AgMERRA and AgCFSR are particularly helpful in that they capture major synoptic events (e.g., heat waves or storm systems) and interannual variability at most locations. These data may then be used to fill in observational gaps following a simple bias-correction procedure, as described for AgMERRA by the following equations:

$$T\text{max}_{\text{estimate}}(d, m) = T\text{max}_{\text{AgMERRA}}(d, m) + \Delta\overline{T\text{max}}_{\text{overlap}}(\overline{m}), \quad (8)$$

$$T\text{min}_{\text{estimate}}(d, m) = T\text{min}_{\text{AgMERRA}}(d, m) + \Delta\overline{T\text{min}}_{\text{overlap}}(\overline{m}), \quad \text{and} \quad (9)$$

$$P_{\text{estimate}}(d, m) = P_{\text{AgMERRA}}(d, m) \times \rho P_{\text{overlap}}(\overline{m}), \quad (10)$$

where Δ terms are determined by examining all days in a given month where AgMERRA and observations exist and then differencing temperatures ($T_{\text{obs}} - T_{\text{AgMERRA}}$), and ρ is a ratio formed in the same manner for precipitation ($P_{\text{obs}}/P_{\text{AgMERRA}}$). Solar radiation and wind data may also be adjusted according to distribution fits or bias ratios, and the relative humidity can be converted to vapor pressure or dewpoint temperature using the revised maximum temperature. The resulting estimate contains the sub-monthly and interannual variability of the AgMERRA dataset while also removing mean

biases between a particular location and the AgMERRA grid box that it corresponds to. It is important to visualize these filled time series data to ensure that the final result matches expectation, and to recognize that in some cases the Δ terms may not be representative of the whole 1980–2010 period (e.g., in situations where AgMERRA T_{max} biases rise in proportion to temperature anomalies). In regions far from the Prime Meridian there is an increasing chance that the local day's weather conditions will differ from the corresponding date in the universal time clock (UTC; pinned to the Prime Meridian) used in AgMERRA and AgCFSR. On occasion it may therefore be necessary to draw T_{max} from a day prior or past the UTC date. In many applications this difference is small compared to the sensitivity of a crop model, and an examination of correlation maps from Figs. 2 and 3, Fig. 5 reveals no substantial longitudinal dependence on daily correlations.

6. Conclusions and future development

The AgMERRA and AgCFSR climate forcing datasets contain the variables required for a large number of agricultural modeling applications on a climate time scale, providing consistent coverage (even in areas where reliable station data are not available) and enhanced resolution of precipitation events in AgMERRA. AgMERRA currently supplies time series for AgMIP's Coordinated Climate Crop Modeling Project (C3MP; Ruane et al., 2014a) and forms the basis of gap-filling for AgMIP's Regional integrated assessments in Sub-Saharan Africa and South Asia (Rosenzweig et al., 2012; Ruane et al., 2014b). AgMERRA and AgCFSR also provide driving datasets for AgMIP's Global Gridded Crop Model Intercomparison (GGCMI; Elliott et al., in review). Owing to its superior performance on sub-seasonal temperature variability, AgCFSR is being used to drive irrigated wheat models for the second phase of the AgMIP Wheat Model Intercomparison (Asseng et al., 2013). These datasets may also be used as an improved historical basis for the generation of future climate scenarios (e.g., Hempel et al., 2013), providing more realistic climate variability and extreme statistics of daily precipitation as a target for statistical downscaling. Both datasets are freely available online (http://data.giss.nasa.gov/impacts/agmipcf), and an AgMIP interface is under development to allow sub-setting and re-formatting of these products for tailored applications.

Future versions of AgMERRA and AgCFSR are under development and will likely confront additional challenges not included in the version presented here. Evolution will be possible with each new release of the gridded observational datasets or reanalysis; a process that has already improved several of these products since this version of AgMERRA and AgCFSR was first calculated. A primary interest is to extend these datasets through at least 2012, which would capture the severe drought conditions experienced that year in the United States. New precipitation products may also hold promise for combination with AgCFSR in a manner similar to the way in which CPCU rainfall provided such benefit to AgMERRA (via MERRA-Land).

Improvement is likely possible in areas with complex terrain, as lapse-rate corrections can improve temperature and relative humidity that have been interpolated onto a grid-box with substantially different mean elevation. For AgMERRA and AgCFSR these corrections will be oriented toward the elevation of agricultural production rather than the mean grid box, which will be important in regions like Greece or Chile where agricultural production occurs in the valleys of tall mountain ranges. Sub-daily-scale versions of these products are also possible and would meet an increasing demand for agricultural applications related to global vegetation and irrigation modeling.

References

Asseng, S., et al., 2013. Uncertainties in simulating wheat yields under climate change. Nat. Clim. Change 3, 827–832.

Adam, J.C., Clark, E.A., Lettenmaier, D.P., Wood, E.F., 2006. Correction of global precipitation for orographic effects. J. Clim. 19, 15–38.

Adam, J.C., Lettenmaier, D.P., 2003. Adjustment of global gridded precipitation for systematic bias. J. Geophys. Res. 108, 4257, http://dx.doi.org/10.1029/2002JD002499.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration—guidelines for computing crop water requirements. FAO Irrig. Drain. Pap. 56 (Rome).

Bassu, S., et al., 2014. How do various maize crop models vary in their responses to climate change factors? Global Change Biol., http://dx.doi.org/10.1111/gcb.12520.

Bosilovich, M.G., Chen, J., Robertson, F.R., Adler, R.F., 2008. Evaluation of global precipitation in reanalyses. J. Appl. Meteorol. Clim. 47, 2279–2299.

Bosilovich, M.G., Robertson, F.R., Chen, J., 2011. Global energy and water budgets in MERRA. J. Clim. 24, 5721–5739, http://dx.doi.org/10.1175/2011JCLI4175.1.

Chen, M., et al., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. 113, D04110, http://dx.doi.org/10.1029/2007JD009132.

Curry, J.A., Webster, P.J., 1999. Thermodynamics of Atmospheres & Oceans. International Geophysics Series, 65. Academic Press, San Diego, CA, pp. 471.

Dee, D.P., et al., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597.

Dunn, R.J.H., et al., 2012. HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. Clim. Past 8, 1649–1679, http://dx.doi.org/10.5194/cp-8-1649-2012.

Dzotsi, K.A., Matyas, C.J., Jones, J.W., Baigorria, G., Hoogenboom, G., 2013. Understanding high resolution space-time variability of rainfall in southwest Georgia, United States. Int. J. Climatol., http://dx.doi.org/10.1002/joc.3904.

Elliott, J., Müller, C., Deryng, D., Büchner, M., Chryssanthacopoulos, J., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, C., Mueller, N., Ramankutty, N., Rosenzweig, C., Ruane, A.C., Sheffield, J., 2014. The Global Gridded Crop Model Intercomparison (GGCMI): data and protocols. Geosci. Model Dev. (in review).

Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2013. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. Int. J. Climatol., http://dx.doi.org/10.1002/joc.3711.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113, D20119, http://dx.doi.org/10.1029/2008JD010201.

Hempel, S., Frieler, K., Warszawski, L., Shewe, J., Piontek, F., 2013. A trend-preserving bias correction—the ISI-MIP approach. Earth Syst. Dyn. 4, 219–236.

Higgins, R.W., Shi, W., Yarosh, E., Joyce, R., 2000. Improved US precipitation quality control system and analysis. In: NCEP/Climate Prediction Center Atlas No. 7. NOAA/NWS, pp. 40.

Hijmans, R.J., et al., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Huffman, G.J., Adler, R.F., Morrissey, M., Bolvin, D.T., Curtis, S., Joyce, R., McGavock, B., Susskind, J., 2001. Global precipitation at one-degree daily resolution from multi-satellite observations. J. Hydrometeor. 2, 36–50.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B., 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeorol. 8, 38–55, http://dx.doi.org/10.1175/JHM560.1.

Hsu, K., Gao, X., Sorooshian, S., Gupta, H.V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteor. 36, 1176–1190.

Iizumi, T., Okada, M., Yokozawa, M., 2014. A meteorological forcing dataset for global crop modeling: development, evaluation, and intercomparison. J. Geophys. Res.: Atmos., http://dx.doi.org/10.1002/2013JD020130.

Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. J. Hydrometeorol. 5, 487–503.

Kalnay, E., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull Amer. Met. Soc. 77, 437–471.

Kanamitsu, M., Ebisuzaki, W., Woolen, J., Potter, J., Fiorino, M., 2002. NCEP-DOE AMIP-II reanalysis (R-2). Bull. Amer. Met. Soc. 83, 1631–1643.

Li, T., et al., 2014. Crop-model ensembles reduce uncertainty in predicting rice yield under climate change. Global Change Biol. (in press).

Lorenz, C., Kunstmann, H., 2012. The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis. J. Hydrometeor. 13 (5), 1397–1420.

Meng, J., Yang, R., Wei, H., Ek, M., Gayno, G., Xie, P., Mitchell, K., 2012. The land surface analysis in the NCEP Climate Forecast System Reanalysis. J. Hydrometeor. 13, 1621–1630, http://dx.doi.org/10.1175/JHM-D-11-090.1.

Monfreda, C., Ramankutty, N., Foley, J.A., 2008. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Global Biogeochem. Cycles 22, GB1022, http://dx.doi.org/10.1029/2007GB002947.

Nelson, G.C., et al., 2013. Climate change effects on agriculture: economic responses to biophysical shocks. Proc. Natl. Acad. Sci., http://dx.doi.org/10.1073/pnas.1222465110.

New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. Clim. Res. 21, http://dx.doi.org/10.3354/cr021001.

Onogi, K., et al., 2007. The JRA-25 reanalysis. J. Meteorol. Soc. Jpn. 85, 369–432, http://dx.doi.org/10.2151/jmsj.85.369.

Reichle, R.H., Koster, R.D., De Lannoy, G.J.M., Forman, B.A., Liu, Q., Mahanama, S.P.P., Touré, A., 2011. Assessment and enhancement of MERRA land surface hydrology estimates. J. Clim. 24, 6322–6338, http://dx.doi.org/10.1175/JCLI-D-10-05033.1.

Reichle, R.H., 2012. The MERRA-land data product (Version 1.2). In: GMAO Office Note No. 3., pp. 38pp, Available at ⟨http://gmao.gsfc.nasa.gov/pubs/office_notes/⟩.

Rienecker, M.R., et al., 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Clim. 24, 3624–3648.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Mutter, C.Z., Adiku, S.G.K., Ahmad, A., Beletse, Y., Gangwar, B., Guntuku, D., Kihara, J., Masikati, P., Paramasivan, P., Rao, K.P.C., Zubair, L., 2012. The Agricultural Model Intercomparison and Improvement Project (AgMIP): integrated regional assessment projects. In: Hillel, D., Rosenzweig, C. (Eds.), Handbook of Climate Change and Agroecosystems: Global and Regional Aspects and Implications (Vol. 2, pp. 263–280). ICP Series on Climate Change Impacts, Adaptation, and Mitigation. Imperial College Press, Potsdam, Germany.

Rosenzweig, C., et al., 2013a. The Agricultural Model Intercomparison and Improvement Project (AgMIP): protocols and pilot studies. Agric. Forest Meteorol. 170, 166–182, http://dx.doi.org/10.1016/j.agrformet.2012.09.011.

Rosenzweig, C., et al., 2013b. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Amer. Acad. Sci., http://dx.doi.org/10.1073/pnas.1222463110.

Rötter, R.P., et al., 2013. Challenges for agro-ecosystem modelling in climate change risk assessment for major European crops and farming systems. In: Proceedings of the Impacts World 2013 Conference at Potsdam, May 2013, Germany, pp. 555–564, http://dx.doi.org/10.2312/pik.2013.001.

Ruane, A.C., Roads, J.O., 2007a. 6-hour to 1-year variance of five global precipitation sets. Earth Interact. 11, http://dx.doi.org/10.1175/EI225.1.

Ruane, A.C., Roads, J.O., 2007b. The diurnal cycle of water and energy over the continental United States from three reanalyses. J. Meteor. Soc. Jpn. 85A, 117–143.

Ruane, A.C., McDermid, S., Rosenzweig, C., Baigorria, G.A., Jones, J.W., Romero, C.C., Cecil, L.D., 2014a. Carbon–temperature–water change analysis for peanut production under climate change: a prototype for the AgMIP Coordinated Climate-Crop Modeling Project (C3MP). Global Change Biol. 20 (2), 394–407.

Ruane, A.C., Winter, J.M., McDermid, S.P., Hudson, N.I., 2014b. AgMIP Climate Datasets and Scenarios for Integrated Assessment. In: Hillel, D., Rosenzweig, C. (Eds.), Handbook of Climate Change and Agroecosystems: The Agricultural

Model Intercomparison and Improvement Project (AgMIP) Phase I Activities (Vol. 3). ICP Series on Climate Change Impacts, Adaptation, and Mitigation. Imperial College Press, Under revision.

Saha, S., et al., 2010. The NCEP climate forecast system reanalysis. Bull. Amer. Meteorol. Soc. 91, 1015–1057.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2011. GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data., http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V6_050.

Schwalm, C.R., Huntinzger, D.N., Cook, R.B., Wei, Y., Baker, I.T., Neilson, R.P., Poulter, B., Caldwell, P., Sun, G., Tian, H.Q., Zeng, N., 2014. A model-data intercomparison of simulated runoff in the contiguous United States: results from the North America Carbon Regional and Continental Interim-Synthesis. Biogeosci. Discuss. 11, 1801–1826, http://dx.doi.org/10.5194/bgd-11-1801-2014.

Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. J. Climate 19, 3088–3111.

Singels, A., Jones, M., Marin, F., Ruane, A.C., Thorburn, P., 2013. Predicting climate change impacts on sugarcane production at sites in Australia, Brazil and South Africa using the Canegro model. Sugar Tech, http://dx.doi.org/10.1007/s12355-013-0274-1.

Stackhouse Jr., P.W., Gupta, S.K., Cox, S.J., Mikovitz, J.C., Zhang, T., Hinkelman, L.M., 2011. The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset. GEWEX News 21 (1), February, 10-12.

Uppala, S., et al., 2005. The ERA-40 re-analysis. Q. J. Roy. Meteorol. Soc. 131, 2961–3012, http://dx.doi.org/10.1256/qj.04.176.

von Lampe, M., et al., 2014. Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison. Agric. Econ. 45 (1), 3–20.

Weedon, G.P., Gomes, S., Viterbo, P., Shuttleworth, W.J., Blyth, E., Österle, H., Adam, J.C., Bellouin, N., Boucher, O., Best, M., 2011. Creation of the WATCH Forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J. Hydrometeorol. 12, 823–848, http://dx.doi.org/10.1175/2011JHM1369.1.

Weedon, G.P., Gomes, S., Balsamo, G., Best, M.J., Bellouin, N., Viterbo, P., 2012. README file for the "WFDEI" dataset. Version: September 18th, 2013, Available at www.eu-watch.org/data_availability

White, J.W., Hoogenboom, G., Stackhouse Jr., P.W., Hoell, J.M., 2008. Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US. Agric. Forest Meteorol. 148, 1574–1584, http://dx.doi.org/10.1016/j.agrformet.2008.05.017.

White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.A., 2011a. Methodologies for simulating impacts of climate change on crop production. Field Crops Res. 124 (3), 357–368.

White, J.W., Hoogenboom, G., Wilkens, P.W., Stackhouse, P.W., Hoel, J.M., 2011b. Evaluation of satellite-based, modeled-derived daily solar radiation data for the Continental United States. Agron. J. 103 (4), 1242–1251, http://dx.doi.org/10.2134/agronj2011.0038.

Wilby, R.L., Charles, S., Zorita, E., Timbal, B., Whetton, P., Mearns, L., 2004. Guidelines for use of Climate Scenarios Developed from Statistical Downscaling Methods. IPCC, IPCC Supporting Material, available from the DDC of IPPC TGCIA.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego, CA, 467pp.

Willmott, C.J., Matsuura, K., 1995. Smart interpolation of annually averaged air temperature in the United States. J. Appl. Meteorol. 34 (12), 2577–2586.

WMO, 1989. Calculation of Monthly and Annual 30-Year Standard Normals (WCDP928 No. 10, WMO-TD/No.341). World Meteorological Organization, Geneva.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., Kitoh, A., 2012. APHRODITE: constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. Bull. Amer. Meteorol. Soc. 93, 1401–1415, http://dx.doi.org/10.1175/BAMS-D-11-00122.1.