SUSTAINABILITY
CONSORTIUM

FOOD, BEVERAGE & AGRICULTURE

Geospatial Climate Data

Ryan Z. Johnston
Marty D. Matlock, Ph.D., P.E., C.S.E.
University of Arkansas
The Sustainability Consortium

White Paper #104 | February 2011

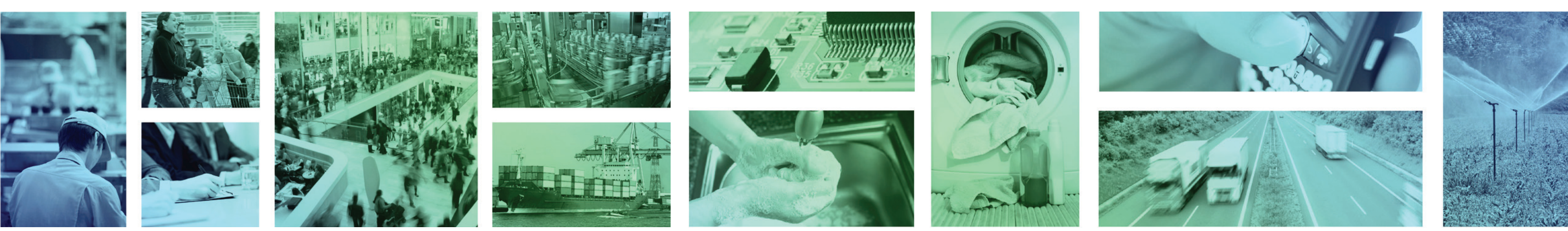

TABLE OF CONTENTS

LIST OF TABLES

1. Introduction

In recent years, plot specific crop models have been adapted to national and regional scales to aid policy makers with agricultural decisions concerning climate change (Mearns, Mavromatis, & Tsvetsinskaya, 1999; Southworth, et al., 2000; Jones & Thornton, 2003; Reilly, et al., 2003; Xiong, Matthews, Holman, Lin, & Xu, 2007) and the resulting effects on food security (Parry, Rosenzweig, Iglesias, Fischer, & Livermore, 1999) and future water demands (Liu, Zehnder, & Yang, 2009;Mo, Liu, Lin, & Guo, 2009). These models are often constrained by data to represent geospatially variable inputs as homogeneous data. The impact of these assumptions on model effectiveness is a function of the sensitivity of the input parameter to the model, the scale of data being aggregated, and the scale of the analysis. The impact of aggregation of geospatially variable data at the regional level is loss of calibration and validation effectiveness and thus utility for most modeling efforts (Hasen & Jones, 2000). Regional cropping systems are highly heterogeneous and model inputs should reflect as much. Considering the vast quantities of available input data, decisions must be made about desired spatial and temporal resolution, as well as the amount of generalizations that can be made about the study in question.

Inputs for the crop modeling process can be separated into three major categories: climatic, soil, and management parameters. The purpose of this document is to review current geospatial datasets available as inputs for crop modeling research at the regional scale. This review will also only focus on regional dataset applicable to the US and globe.

2. Geospatial Climate Data

Historical climate data that is geospatially explicit is a necessity for any crop modeling process, especially on a regional scale. Crop simulation models typically require large amounts of climatic input data, including maximum and minimum temperature, precipitation, and solar radiation at a daily time step. This information can be provided by a variety of datasets, each with its own advantages and disadvantages. A few publically available sources include weather station data, interpolated grids based on station data, and satellite derived data. Depending on the application, the user must select the appropriate dataset. The following provides an overview of some of the more popular sources of historical climate datasets

3. Weather Stations

Weather station data is one of the most used sources of historical meteorological/climate data available. Depending on the study, anything from local weather stations to a global network of stations can be used as inputs for crop modeling.

A. NCDC

One of the most well known sources of regional weather station data comes from National Climatic Data Center (NCDC) sector of the National Oceanic and Atmospheric Administration (NOAA). Daily observations of temperature, precipitation, winds, pressure, snow, and others can be found for over 15,000 stations worldwide, including over 2,000 stations in the United States from the Global Surface Summary of the Day (GSOD). The data covers a temporal range that begins in 1929 to near present time, although the most complete records begin in 1973. While this dataset represents one of the most comprehensive, it suffers from data gaps in localities of low to no representation. Data access can found at the following link:

http://www7.ncdc.noaa.gov/CDO/cdoselect.cmd?datasetabbv=GSOD&countryabbv=&georegionabbv=

An alternative to the Global Summary of the Day weather station database is the Global Historical Climate Network Data (GHCN) version 2. The GHCN provides historical temperature, precipitation, and pressure data for thousands of land stations on a monthly basis. Precipitation is represented over 20590 stations, mean temperature over 7280 stations, and minimum and maximum temperature over 4966 stations. This data also suffers large data gaps favoring the Northern Hemisphere.

http://www.ncdc.noaa.gov/ghcnm/v2.php

B. FAOCLIM 2.0

Other sources of weather station data include the FAOCLIM 2.0 global climate database. FAOCLIM 2.0 contains monthly data for weather stations across the world. This station database contains a number of variables including a monthly total of evapotranspiration, precipitation, and sunshine duration, and monthly mean values for maximum and minimum temperature, vapor pressure, and wind speed. FAOCLIM 2.0 also contains both long-term averages (1961-1990) and time series for precipitation and temperature.

http://www.fao.org/nr/climpag/pub/EN1102_en.asp

Weather station data can be a valuable in the crop modeling process. Often, local weather station data is obtained for a particular farm and assumed to be representative of the weather conditions occurring on site. Quality weather station data is invaluable for farm based studies; however, regional studies require historical climate variables to be manipulated into a format suitable for large area. Common techniques include creating Thiessen Polygons or interpolating station data into grids.

C. Gridded Climatic Datasets

While deriving Thiessen Polygons from station data is a common practice in hydrology, regional crop modeling studies typically use gridded climatic datasets. Gridded datasets offer the advantages of easily being integrated into Geographical Interface Systems (GIS), undergone quality assurance protocols at the station level, and efficiently convey a vast amount of information compared to the colossal weather station archives. A variety of gridded datasets exist at different spatial and temporal resolutions.

D. PRISM (Parameter-Elevation Regressions on Independent Slopes Model)

The PRISM (Parameter-Elevation Regressions on Independent Slopes Model) climate mapping system is a regression-based model that uses point measurements of precipitation, temperature, and other climactic factors, a digital elevation model (DEM), other spatial data sets, and human-expert knowledge to generate digital grid estimates of monthly climatic parameters (Daly et al. 2002). The PRISM dataset contains a very high resolution (4-km) of monthly estimates of mean, maximum and minimum temperature beginning in 1895 and extends near present time (Table 3-1).

Table 3-1. PRISM Climate Data

Available Parameters	Temporal Coverage	Temporal Resolution	Spatial Coverage	Spatial Resolution
Mean Temperature	1895-Near Present Time	Monthly	Conterminous Untied States	4-km
Maximum Temperature	1895-Near Present Time	Monthly	Conterminous Untied States	4-km
Minimum Temperature	1895-Near Present Time	Monthly	Conterminous Untied States	4-km
Precipitation	1895-Near Present Time	Monthly	Conterminous Untied States	4-km

PRISM takes a unique approach to interpolating historical weather data. PRISM relies on the assumption that for a localized region, elevation is the most important factor in the distribution of many climate elements, such as temperature and precipitation. A linear regression function, between climate and elevation, serves as the main predictive equation when estimating grid values from station data. The equation takes the following form (Daly et al. 2002):

$$Y = \beta_1 X + \beta_0;\ \beta_{1m} \leq \beta_1 \leq \beta_{1x}$$

where Y is the predicted climate element (temperature or precipitation), $\beta_1$and β_0 are the regression slope and intercept, X is the DEM elevation at the target grid cell, and $\beta_{1m}$and $\beta_{1x}$are the minimum and maximum allowable regression slopes (see Daly et al. 2002 for a more detailed explanation of boundary values).

In addition to elevation, input weather station data are assigned additional weights based on an expert knowledge base of key climatic forcing factors. Key factors include distance from target cell, elevational influence on climate, terrain-induced climate transitions, coastal proximity, two-layer atmospheric effects, and orographic effectiveness of terrain (Daly et al. 2002). The end result is a very high resolution grid of climate variables.   Data can be found at the following link: http://www.prism.oregonstate.edu/docs/index.phtml

While PRISM has produced many high resolution climate grids that cover over 100 years of observations, the temporal resolution is lacking. Values for climate variable are only available on a monthly basis. To be used in the crop modeling process, a weather generator would be needed.

E. WorldClim

An alternative to the very high resolution dataset PRISM is WorldClim, which has been developed at an even high resolution for the entire globe, excluding Antarctica. WorldClim contains global estimates of monthly mean, maximum, and minimum temperature and precipitation at a 1-km resolution between the years 1950-2000.

Similar to PRISM, WorldClim interpolated weather station to produce monthly grids of climate variables. Weather station data came from a variety of different weather station databases including GHCN, WMO climatological normals, FAOCLIM 2.0, International Center for Tropical Agriculture (CIAT), and other regional databases. Extensive quality control measures were taken to insure no duplicate records were present after combining the various databases, giving precedence to the GHCN database. After the quality control check, the database consisted of precipitation records from 47554 locations, mean temperature from 24542 locations, and minimum and maximum temperatures from 14835 locations (Hijmans et al. 2005).

Once the weather stations were checked, the ANUSPLIN software package version 4.3 was used to interpolate global climate surfaces from the weather stations. This software implements thin-plate smoothing spline procedure, using every station as a data point. The authors fitted a second-order spline using latitude, longitude, and elevation as independent variables, which produced the lowest overall cross –validation errors (Hijmans et al. 2005). Considering the ANUSPLIN program creates a continuous surface projection, the LAPGRD program was to create a global grid. Hence, the end resolution of the global grid merely depends on the input grid; the higher the resolution of the input grid, the better it represents the modeled climate data (Hijmans et al. 2005). The end result was a very high resolution dataset (1-km).

While this dataset represents an impressive display, the authors did express concern of the resulting precipitation grids. The WorldClim precipitation grids compared very well with other datasets, including PRISM, in the lower elevations of the eastern United States. However, major differences were observed in the higher elevations of the Rocky Mountains of the western United States. The authors concluded that one cannot be very certain about the values of any particular grid cell in mountains regions (Hijmans et al. 2005). In addition, the major differences were seen when compared the gridded dataset of New et al. 2002. The differences are geographically related to areas with a low density of weather stations, such as Greenland, and remote parts of Africa and South America (Hijmans et al. 2005). It is unclear how these discrepancies would affect crop modeling studies. The WorldClim data can be found at the following link:

http://www.worldclim.org/

F. VEMAP (Vegetation/Ecosystem Modeling and Analysis Project)

The Vegetation/Ecosystem Modeling and Analysis Project (VEMAP) provides a public database that includes high resolution climate and soils data of the conterminous USA on a 0.5° grid (Wu et al 2010)(Table 3-2). The project began as a multi-agency, international program to simulate and understand ecosystem dynamics for the continental United States to altered climate and elevated atmospheric CO_2 concentrations (Kittel et al 1995). It continues to represent a source of climate and soil data used in a variety of scientific topics related to crop modeling (Jagtap & Jones, 2002; Irmark, Jones, & Jagtap, 2005; Wu, Liu, Hoogenboom, & White, 2010).

Table 3-2. WorldClim Climate Data				
Available Parameters	Temporal Coverage	Temporal Resolution	Spatial Coverage	Spatial Resolution
Mean Temperature	1950-2000	Monthly	Global	1-km
Maximum Temperature	1950-2000	Monthly	Global	1-km
Minimum Temperature	1950-2000	Monthly	Global	1-km
Precipitation	1950-2000	Monthly	Global	1-km

VEMAP includes daily minimum and maximum temperature, precipitation, and solar radiation from 1895 to 1993 for each grid cell. Each of these values was determined in a unique way. Temperature values were derived from monthly mean minimum and maximum observations from 4613 weather station normals provided by NCDC. To account for the effect of topography on temperature, each of the provided normals were adiabatically adjusted to sea level using algorithms developed by Marks and Dozier (1992). The adjusted temperatures were then interpolated to the corresponding VEMAP 0.5° grid cell and then adiabatically readjusted to grid elevations (VEMAP Members 1995). Precipitation values were derived from a previously developed dataset, the Precipitation-Elevation Regressions on Independent Slopes Model. These grids were aggregated to 0.5° resolution. To complete the VEMAP dataset with daily values of the climatic parameters, daily weather generators were used. A modified version of WGEN was used to simulate daily values of temperature and precipitation. Climate Simulator (CLIMSIM), a simplified version of Mountain Microclimate Simulator (MT-CLIM), was used to estimate daily total incident solar radiation, daily irradiance and surface humidity based on the daily minimum and maximum temperature and precipitation (VEMAP Members 1995).

The VEMAP database offers one additional parameter that sets it apart from other sources of historical climate information. VEMAP also contains soil information for each grid. Measurements include bulk density, sand, silt, clay, organic content, and rock fragment for up to four dominant categories. These values were based on Kern's (1994, 1995) 10-km gridded Soil Conservation Service national-level (NATSGO) database. A cluster analysis was used to group the 10-km sub-gird elements into modal soil types. Cell soil properties are represented by a set of 1-4 modal soil profiles rather than by an average that may not correspond to an actual soil in the region (VEMAP Members 1995). http://www.cgd.ucar.edu/vemap/datasets.html

G. Climate Research Unit (CRU)

The Climate Research Unit (CRU) has compiled numerous datasets in reference to natural and anthropogenic climate change. These datasets are well known and have been the subject of much criticism following a massive email highjack in 2009. Regardless, the data represents one of the premier climatology databases and is used throughout scientific community.

The CRU database contains two different types, one at a course resolution (above 2°) and one at a fine resolution (0.5°). There are two datasets that make up the coarse resolution data. One set is the HadCRUT3 for land surface temperature. This set represents over 4349 weather stations and contains land air temperature anomalies on a 5° X 5° grid dating back to 1850. This dataset contains many improvements to older versions (HadCRUT and HadCRUT2), including more station data to improve global coverage, removal of duplicate stations, and improving station normals and standard deviations (Brohan et al. 2006). Precipitation data is contained in the other and can be found at a 2.5° X 3.75° resolution. This set extends from 1900 to 1998 (Hulme 1998). However, while the gridded precipitation data is at a higher resolution comparatively, spatial coverage is spotty; leaving major gaps in northern North America, Central Asia, and Africa. CRU however holds a high resolution dataset that is of interest for crop modeling.

The CRU TS 3.0 dataset contains monthly averages of six climate elements including precipitation, mean, maximum, and minimum temperature, and other (Table 3-3). This dataset represents an update to the previous versions (CRU TS 1.0, CRU TS 2.0, and CRU TS 2.1) and contains the highest temporal and parameter coverage. Weather station records were obtained from a variety of previously compiled sources and checked for inhomogeneities using an approach similar to the GHCN automatic method of homogenization (Mitchell and Jones 2005). This method uses neighboring stations to construct a reference series for which to compare candidate station data. Once reference series were constructed and inhomogeneities were corrected, station data were merged into one database. This database was then converted to anomalies relative to the 1961-1990 normal. Climate anomalies were then interpolated onto a continuous surface and a global 0.5° grid was derived (Mitchell and Jones 2005). This data can be found at the following link: http://www.cru.uea.ac.uk/cru/data/hrg/

While this dataset is at a high resolution and a global spatial coverage, the data is lacking temporal resolution necessary for crop modeling. In order to use this CRU TS 3.0 data, a weather generator must be employed to create a daily time series of climate observations. Candidate generators include dGen-CRU, Weatherman, and WGEN.

Table 3-3. VEMAP Climate Data

Available Parameters	Temporal Coverage	Temporal Resolution	Spatial Coverage	Spatial Resolution
Mean Temperature	1895-1993	Daily	Conterminous US	0.5°X0.5°
Minimum Temperature	1895-1993	Daily	Conterminous US	0.5°X0.5°
Maximum Temperature	1895-1993	Daily	Conterminous US	0.5°X0.5°
Precipitation	1895-1993	Daily	Conterminous US	0.5°X0.5°
Solar Radiation	1895-1993	Daily	Conterminous US	0.5°X0.5°
Wind Speed	1895-1993	Monthly	Conterminous US	0.5°X0.5°

H. NASA POWER (Prediction of Worldwide Energy Resource) Agroclimatology Data

One source of historical meteorological data is the NASA Agroclimatology Archive, one component of NASA's POWER (Prediction of Worldwide Energy Resource) project. POWER was created to allow access to data derived from NASA's Surface Meteorological and Solar Energy (SSE) project for those interested in the design of renewable energy systems. The Agroclimatology archive was developed with agricultural Decision Supports Systems (DSS) in mind and provides easy download of historical data for specific site locations. The parameters contained in this dataset are based upon solar radiation derived from satellite observations and meteorological data from the Goddard Earth Observing System assimilation model. The archive boasts globally comprehensive coverage at 1° latitude by 1° longitude grid dating back to July 1983 to near present time (Table 3-4). Reported parameters include, top-of-atmosphere insolation, insolation on a horizontal surface, downward long radiative flux, daily mean, maximum, and minimum temperatures at 2m above ground surface, relative humidity at 2m above ground surface, dew point at 2m aboveground surface, wind spend at 10m above ground surface and precipitation (starts January 1997 and ends August 2009). Access to the data can be found at the following website:

http://earth-www.larc.nasa.gov/cgi-bin/cgiwrap/solar/agro.cgi?email=agroclim@larc.nasa.gov

The values of each parameter come from a combination of different sources. Solar radiation values were either obtained from or derived from parameters in the NASA/GEWEX Surface Radiation Budget (SRB) 3.0 archive. Radiation parameters are based on a globally comprehensive Earth energy budget. Meteorological parameters were obtained from NASA's Global Model and Assimilation Office (GMAO), Goddard Earth Observing System global assimilation model version 4 (GEOS-4). GEOS-4 provides global estimates of atmospheric parameters, such as temperature, on a 3 hour time step. The original GEOS-4 parameters were outputted on a 1° X 1.25° grid and were bi-linearly interpolated for the SSE project. Precipitation data was obtained from the Global Precipitation Climate Project (GPCP) and the Tropical Rainfall Measurement Mission (TRMM) Daily Global and Regional Rainfall derived data sets. The TRMM data are used to fill in gaps within the GPCP data between latitudes of 40°N and 40°S. Finally, wind speed values are based on GEOS-1 data with few adjustments being made due to updates in vegetative areas and new science information (NASA 2010).

Table 3-4. CRU TS 3.0 Climate Data

Available Parameters	Temporal Coverage	Temporal Resolution	Spatial Coverage	Spatial Resolution
Cloud Cover	1901-2006	Monthly	Global	0.5° X 0.5°
Diurnal Temperature Range	1901-2006	Monthly	Global	0.5° X 0.5°
Frost Day Frequency	1901-2006	Monthly	Global	0.5° X 0.5°
Precipitation	1901-2006	Monthly	Global	0.5° X 0.5°
Daily Mean Temperature	1901-2006	Monthly	Global	0.5° X 0.5°
Monthly Average Daily Minimum Temperature	1901-2006	Monthly	Global	0.5° X 0.5°
Monthly Average Daily Maximum Temperature	1901-2006	Monthly	Global	0.5° X 0.5°
Vapor Pressure	1901-2006	Monthly	Global	0.5° X 0.5°
Wet Day Frequency	1901-2006	Monthly	Global	0.5° X 0.5°

The accuracy of the POWER dataset was compared to ground site data on a global basis. It is generally accepted that quality ground-measured data are more accurate than satellite-derived values. However, there do exist major sources of uncertainties associated with ground based measurements including calibration drift, operational uncertainties, or data gaps that are unknown or unreported for most ground site data sets (NASA 2010). Regardless, radiation parameters were compared to data from the Baseline Surface Radiation Network (BSRN), meteorological parameters were compared to data from the National Climate Data Center (NCDC), and wind speeds were compared to RETScreen Weather Database (NETScreen 2005). The following tables were obtained from NASA (2010) and provide summary statistics of the comparisons (Table 3-5 through Table 3-8)(NASA 2010).

Table 3-5. NASA POWER Agroclimatology Archive				
Available Parameters	Temporal Coverage	Temporal Resolution	Spatial Coverage	Spatial Resolution
Top-of-Atmosphere Insolation	July 1983 – near present	Daily	Global	1° X 1°
Insolation on Horizontal Surface	July 1983 – near present	Daily	Global	1° X 1°
Downward Longwave Radiative Flux	July 1983 – near present	Daily	Global	1° X 1°
Mean Temperature (°C)	January 1983 – near present	Daily	Global	1° X 1°
Maximum Temperature (°C)	January 1983 – near present	Daily	Global	1° X 1°
Minimum Temperature (°C)	January 1983 – near present	Daily	Global	1° X 1°
Relative Humidity	January 1983 – near present	Daily	Global	1° X 1°
Dew Point	January 1983 – near present	Daily	Global	1° X 1°
Precipitation	January 1997 – August 2009	Daily	Global	1° X 1°

Table 3-6. Regression analysis of SSE versus BSRN monthly averaged values for the time period July 1983 through June 2006 (NASA 2010)			
Parameter	Region	Bias (%)	RMSE (%)
Horizontal Insolation	Global 60° Poleward 60° Equatorward	-2.55 -8.44 -1.45	13.50 32.19 10.30
Horizontal Diffuse Radiation	Global 60° Poleward 60° Equatorward	7.49 11.29 6.86	29.34 54.14 22.78
Direct Normal Radiation	Global 60° Poleward 60° Equatorward	-4.06 -15.66 2.40	22.73 33.12 20.93

Table 3-7. Linear least squares regression analysis of SSE versus NCDC monthly averaged values for the time period 1983 through 2006 (NASA 2010)					
Parameter	Slope	Intercept	R2	RMSE	Bias
Tmax (°C)	0.99	-1.58	0.95	3.12	-1.83
Tmin (°C)	1.02	0.10	0.95	2.46	0.24
Tavg (°C)	1.02	-0.78	0.96	2.13	-0.58
Tdew (°C)	0.96	-0.80	0.95	2.46	-1.07
RH (%)	0.79	12.72	0.56	9.40	-1.92
Heating Degree Days (degree days)	1.02	12.47	0.93	77.20	17.28
Cooling Degree Days (degree days)	0.86	2.36	0.92	28.90	-5.65
Atmospheric Pressure (hPa)	0.89	102.16	0.74	27.33	-10.20

Table 3-8. Estimated uncertainty for monthly averaged wind speed for the time period July 1983 through June 1993 (NASA 2010)			
Parameter	Method	Bias	RMSE
Wind Speed at 10 meters for terrain similar to airports (m/s)	RETScreen Weather Database (documented 10-m height airport sites) RETScreen Weather Database (unknown-height airport sites)	-0.2 -0.0	1.3 1.3

4. Soils

Considering how heavily most crop models are based on soil-plant-atmosphere interactions, soils are an important input parameter and require the most accurate information available. Various regional datasets are available for different regions throughout the world; although a majority of regional studies use soil profiles supplied by their regions soil survey office.

A. STATSGO2

This soil database is offered by National Resource Conservation Service and is a second addition to the STATSGO dataset. Created in 2006, STATSGO2 provides regionally focused dataset for the conterminous US. STATSGO data has been used in variety of US related studies including (Quiring & Legates, 2008; Grassini, Yang, & Cassman, 2009). STATSGO was created by generalizing more detailed soil survey maps. STATSGO can be used to assign important soil parameters for crop models, including soil classification, color, slope, drainage class, number of soil layers and depth of layer. Many of these variables are very important to the crop modeling process.

http://soildatamart.nrcs.usda.gov

B. ISRIC-WISE

ISRIC-WISE soils database is a globally comprehensive dataset at one of the highest resolutions available at a 5’ X 5’ resolution. The data were created using the soil distribution show on the 1:5 million scale FAO-Unesco Soil Map of the World (DSMW) and soil parameter estimates derived from ISRIC’s global WISE soil profile database (Batjes, 2006). The dataset contains information on 19 soil variables that are commonly used in crop modeling. This dataset has be used by Mekonnen & Hoekstra (2010) the their assessemnt of global production of wheat's green, blue and grey water footprint.

http://www.isric.org/UK/About+Soils/Soil+data/Geographic+data/Global/WISE5by5minutes.htm

5. Regional Management Data

At regional to global scales, crop management data is one of the hardest input parameters to define. Management data must take into account several areas of management including planting and harvesting dates, irrigation amount and timing, fertilizer application amount and timing, and various other strategies. The difficulty arises when one begins to consider the variability in cropping techniques at a regional scale. Take for example any county in the US. Although farms found with the county will experience a similar set of climatic interactions and mostly likely contain a similar set of soil properties, management practices can vary drastically from one farm to the next. This makes accounting for management variability at the regional level nearly impossible and fosters the need for generalizations for any regional modeling process.

The success or failure of such an endeavor relies on finding a fine balance between oversimplifying the input parameters and creating an excessively complex set of inputs. While climate and soil data is currently available in gridded datasets, management practices are not. The following describes a few current techniques used be the scientific community in dealing with this issue.

A. In field cropping techniques

Describing infield cropping techniques, such as planting dates, row spacing, planting density, etc, is a tremendous chore that requires generalizations at the regional scale. Quiring and Legates (2008) used field trial reports collected by the Delaware Cooperative Extension Service to define management parameters such as row spacing, seeding density, seeding depth, date of planting, date of harvest, etc., in a within-season maize prediction study for the state of Delaware. In an assessment of residue retention in maize cropping systems across the state of Jalisco, Mexico, Hartkamp et al. (2004) used expert knowledge and literature sources to determine plant densities, dates, residue and fertilizer amounts. Mo et al. (2009) used local knowledge to assign planting dates for a double cropping system of wheat and corn predict regional crop yield, water consumption and water use efficiency with respect to climate change in the North China Plain. Whether infield cropping techniques are assigned from local extension services or a regional knowledge base, generalization must be made.

B. Fertilizer

Fertilizer application is a major component of crop modeling; however the amount and type of application is hard to identify at the regional scale. One approach is to derive fertilizer volume applied for a crop from the FAO country based annual social-economic database (Tan and Shitbasaki 2003). Similarly, Liu et al. (2007) used FAOSTAT to approximate fertilizer consumption on a global scale. Here, a country's total fertilizer consumption was divided by its total hectares of arable and permanent cropland. Fertilizer application was assumed to the uniform across a country. Similarly to infield cropping practices, modelers can use expert knowledge on nitrogen application rate. The last approach is to assume nitrogen is non-limiting within the system. Several authors have taken this approach when evaluating the affects of climate change (Xiong, Matthews, Holman, Lin, & Xu, 2007; Lal, Singh, Rathore, Srinivasan, & Saseedran, 1998).

C. Irrigation

Much to the same degree as fertilizer application, irrigation at the regional scale is difficult to describe and attempts are derived from a digital global map of irrigated areas generated by the Center for Environmental Systems Research, University of Kassel. This map contains the total percentage that is equipped for irrigation on a gridded basis at a resolution of 30min X 30min. The map can be used in combination with AQUASTAT, a database provided by FAO. AQUASTAT contain data for agricultural water withdrawal and irrigation efficiency. The volume of water applied can then calculated by multiplying water withdrawal and water use efficiency (Liu et al. 2007). Few other studies have attempted to define irrigation volumes beyond this method and irrigation is usually set to off or automatic setting, depending on whether the region of interest falls under rainfed conditions.

6. Conclusion

Geospatial cropping data are an important input for any crop modeling process at the regional scale. Many different sources of such data are available and it is up to the modeler to determine which is best for their particular study. In addition, the model must walk a fine line between setting geospatial parameters and generalizations.

A variety of geospatial climate data exist for use in regional cropping analysis. For predictive studies on farm yields, local weather station data should suffice. However, as the studies are increases, regional grids of interpolated climate observations are needed. PRSIM provides a very high resolution dataset of the conterminous U.S., but lacks the daily resolution needed for crop modeling. An alternative to PRSIM is WorldClim, which produced an even higher resolution dataset for the entire globe (excluding Antarctica). WorldClim also suffered from low temporal resolution and also lacked predictive capabilities in mountainous areas. CRU also offers a relatively high resolution dataset of the entire global, but also suffers from a lack of daily values. A weather generator would be needed for either of these databases before they could be used in a crop modeling study. Other databases do offer daily estimates of climate variables, including VEMAP and NASA POWER. VEMAP provides daily observations of temperature and precipitation (precipitation is derived from the PRISM database) that date back to 1895 both only extends to 1993. Compared the for mentioned datasets, the NASA POWER Agroclimatology database is unique in that is not based on weather station data, rather derived from satellite observations. This database offers daily estimates of temperature, precipitation and solar radiation; however, the dataset is much coarser (1°X 1°).

Soils data is valuable input for any crop modeling process; however, few globally explicit geo-referenced exist. The two most up-to-date datasets are the STATSGO2 data set, pertaining to the conterminous US, and ISRIC-WISE dataset, which is globally comprehensive.   Both dataset contain information necessary for crop modeling.

Finally, crop management data is a necessity for the crop modeling process, although describing management strategies at the regional or global scale is impossible without generalizations. Several approaches have been used to describe the various aspects of management strategies. Information from local extension offices or expert knowledge can be applied to regional crop simulations when estimating infield cropping practices. Typical nitrogen applications can be applied to a crop simulation at the regional scale. They can also be derived from FAOSTAT, with the assumption of uniform country distribution. Finally, irrigation, much to the same effect as nitrogen application, is a vital part of the simulation process. However, regional or global datasets describing irrigation are few and far between. Country based assumptions can be made relating to irrigation following the procedure used by Liu et al. (2007). In the end, it is up to the modeler to select the best input data for the simulation of interest. Decisions must be made concerning the degree of spatial and temporal resolution as well as acceptable generalizations at the regional scale.

7. References

Batjes, N. H. (2006). *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid. Report 2006/02.* Wageningen: ISRIC-WISE derived soil properties on a 5 by 5 arcminutes.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research, 111*.

Chen, J., Liu, H., Garcia, Garcia, A., & Hoogenboom, G. (2010, January). Parameterizing soil and weather inputs for crops simulation models using the VEMAP database. *Agriculture, Ecosystems & Environment, 135*(1-2), 111-118.

Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L., & Pasteris, P. (2002). A knowledge-based approach to the statistical mapping of climate. *Climate Research, 22*, 99-113.

Grassini, P., Yang, H., & Cassman, K. G. (2009). Limits to maize productivity in Western Corn-Belt: A simulation analysis for fully irrigated and rainfed conditions. *Agriculture and Forest Meteorology, 149*, 1254-1265.

Hartkamp, A., White, J., Rossing, W., Ittersum, M. v., Bakker, E., & Rabbinge, R. (2004). Regional application of a cropping systems simulation model: crop residue retention in maize production systems of Jalisco, Mexico. *Agricultural Systems, 82*, 117-138.

Hasen, J. W., & Jones, J. W. (2000). Scaling-up crop models for climate variablity applications. *Agricultural Systems, 65*, 43-72.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005, December). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965-1978.

Hulme, M., & Osborn, T. J. (1998, September). Precipitation sensitivity to global warming: Comparison of observations with HadCM2 simluations. *Geophysical Research Letters, 25*(17), 3379-3382.

Irmark, A., Jones, J. W., & Jagtap, S. S. (2005). Evaluation of the CROPGRO-soybean model for assessing climate impacts on regional soybean yields. *Transactions of the ASAE, 48*(6), 2343-4353.

Jagtap, S. S., & Jones, J. W. (2002). Adaptation and evaluation of the CROPGRO-soybean model to predict regional yield and production. *Agriculture, Ecosystem and Environment, 93*, 73-85.

Jones, P. G., & Thornton, P. K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change, 13*, 51-59.

Kittel, T. G., Rosenbloom, N. A., Painter, T. H., Schimel, D. S., & Participants, V. M. (1995). The VEMAP intergrated database for modelling United States ecosystem/vegetation sensitivity to climate change. *Journal of Biogeography, 22*(4/5), 857-862.

Lal, M., Singh, K., Rathore, L., Srinivasan, G., & Saseedran, S. (1998). Vulnerability of rice and wheat yields in NW India to future changes in climate. *Agricultural and Forest Meteorology, 89*, 101-114.

Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H. (2007). GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scale. *Agricutlural Systems, 94*, 478-493.

Liu, J., Zehnder, A. J., & Yang, H. (2009, May 28). Global consumptive water use for crop production: the importance of green aand virtual water. *Water Resources Research, 45*(5), W05428.1-W05428.15.

Marks, D., & Dozier, J. (1992, November). Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada 2. Snow cover energy balance. *Water Resources Research, 28*(11), 3043-3054.

Mearns, L. O., Mavromatis, T., & Tsvetsinskaya, E. (1999, March 27). Comparative responses of EPIC and CERES crop models to high and low spatial resolution climate change scenarios. *Journal of Geophysical Research, 104*(D6), 6623-6646.

Mekonnen, M. M., & Hoekstra, A. Y. (2010). A global and high-resolution assessment of the green, blue and grey water footprint of wheat. *Hydrology and Earth System Sciences, 14*, 1259-1276.

Members, V. (1995, December). Vegetation/ecosystem modeling and analysis project: Comparing biogegraphy and biogeochemistry models in a continental-scale study of terrestial ecosystem responses to climate change and CO2 doubling. *Global Biogeochemical Cycles, 9*(4), 407-437.

Mitchell, T. D., & Jones, P. D. (2005, May). An improve method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology, 25*(6), 693-712.

Mo, X., Liu, S., Lin, Z., & Guo, R. (2009). Regional crop yield, water consumption and water use efficiency and their responses to climate change in the North China Plain. *Agriculture, Ecosystems and Environment, 134*, 67-78.

NASA POWER Team. (2010). *Sufarce meteorology and solar energy (SSE) release 6.0 methodology VERSION 2.4.1.* NASA POWER Agroclimatology, Retrived from http://power.larc.nasa.gov/documents/SSE_Methodology.pdf.

Parry, M., Rosenzweig, C., Iglesias, A., Fischer, G., & Livermore, M. (1999). Climate change and world food security: a new assessment. *Global Environmental Change, 9*, S51-S67.

Quiring, S. M., & Legates, D. R. (2008). Application of CERES-Maize for within-season prediction of rainfed corn yields in Delaware, USA. *Agricutlural and Forest Meteorology, 148*, 964-975.

Reilly, J., Tubiello, F., McCarl, B., Abler, D., Darwin, R., Fuglie, K., et al. (2003). U.S. agriculture and climate change: new results. *Climatic Change, 57*, 43-69.

Southworth, J., Randolph, J. C., Habeck, M., Doering, O. C., Pfeifer, R. A., Rao, D. G., et al. (2000). Consequences of future climate change and changing climate variability on maize yields in the midwestern Unitied States. *Agriculture, Ecosystems and Environment, 82*, 139-158.

Tan, G., & Shibasaki, R. (2003). Global estimation of crop productivity and the impacts of global warming by GIS and EPIC integration. *Ecological Modelling, 168*, 357-370.

Wu, W., Liu, H.-B., Hoogenboom, G., & White, J. W. (2010). Evaluating the accuracy of VEMAP daily weather data for application in crop simulations on a regional scale. *European Journal of Agronomy, 32*, 187-194.

Xiong, W., Matthews, R., Holman, I., Lin, E., & Xu, Y. (2007). Modelling China's potential maize production at regional scale under climate change. *Climatic Change, 85*, 433-451.