5-2017

Bias and Other Error in Gridded Weather Data Sets and Their Impacts on Estimating Reference Evapotranspiration

Philip A. Blankenau
University of Nebraska-Lincoln, pblankenau@msn.com

BIAS AND OTHER ERROR IN GRIDDED WEATHER DATA SETS AND THEIR IMPACTS ON ESTIMATING REFERENCE EVAPOTRANSPIRATION

by

Philip A. Blankenau

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Civil Engineering

Under the Supervision of Professor Ayse Kilic

Lincoln, Nebraska

May, 2017

BIAS AND OTHER ERROR IN GRIDDED WEATHER DATA SETS AND THEIR IMPACTS ON ESTIMATING REFERENCE EVAPOTRANSPIRATION

Philip A. Blankenau, M.S.

University of Nebraska, 2017

Advisor: Ayse Kilic

Gridded weather data sets are increasingly used in a variety of hydrologic and agricultural applications due to their complete spatial and temporal coverage. One application of gridded data sets is the estimation of evapotranspiration (ET). Several operational remote sensing (RS) approaches for estimating ET, such as the SEBAL, METRIC and EEFlux models, require estimates of reference ET (ET_{ref}), where ET_{ref} is expected ET from a hypothetical reference crop of clipped grass or alfalfa. Gridded weather data provide for the computation of ET_{ref} in all areas of a remote sensing image, and therefore potentially remove the need for dense weather station data.

Given the increasing use of gridded weather data to estimate ET_{ref}, this study assessed the quality of gridded weather data estimates of ET_{ref}. To accomplish this evaluation, several gridded weather data sets – GLDAS-1, NLDAS-2, CFSv2 operational analysis, GRIDMET, RTMA and NDFD – were compared to weather station data that were considered to represent 'ground truth' across the continental United States. The stations were selected to represent reference conditions when possible. The four primary weather variables – near-surface air temperature, vapor pressure, wind speed and shortwave solar radiation - required to compute ET_{ref}, plus calculated ET_{ref} itself were compared.

The application of the same analysis to multiple gridded data sets made comparisons among the gridded data sets possible. Generally, the gridded weather data sets overestimated ET_{ref}. This was mainly due to overestimation of air temperature, shortwave radiation and wind speed, and underestimation of vapor pressure. RTMA had the most accurate weather data and the most accurate estimates of ET_{ref} due to its assimilation of vast amounts of surface weather data and its continual refinement. Surprisingly, the global data sets, GLDAS and CFSv2, generally performed better than their North American counterparts – NLDAS and GRIDMET. All gridded weather data sets may be useful for estimating ET_{ref} and employment in remote sensing ET models provided some procedures for correcting biases are developed.

ACKNOWLEDGEMENTS

I am indebted to many people for the successful completion of this thesis. First, I would like to thank my adviser Dr. Ayse Kilic for the research expertise and support she provided along the way. I would also like to thank Dr. Richard Allen who provided much needed advice at crucial decision points. Other members of my committee who devoted time to reading my preliminary findings and providing fresh ideas also have my appreciation.

I am grateful to the members of my research group who helped download data and set me on the right path to learning programming. The many people I have been in correspondence with regarding weather networks and gridded data sets have also been invaluable, and I thank them for their time.

Lastly, special thanks to my friends and family, and especially to my parents and my girlfriend, Rebecca Nelson, for providing a strong support system for me during my graduate studies.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Gridded weather data sets are increasingly used in a variety of hydrologic, ecologic and agricultural modeling applications. Weather data for modeling have historically been gathered from meteorological weather stations. Weather stations, as a data source, offer several challenges: they may not be close enough to the particular area being modeled to sufficiently represent the local weather; in models applied over large areas their density may not capture the spatial variability of weather within the area; and point data sets often contain periods of missing data or have limited periods of record, creating challenges for models representing long timespans. For these reasons, gridded weather data can provide a potential substitute for ground-based weather measurements.

Gridded weather data have begun to be used for modeling of evapotranspiration (ET), because they provide the necessary information to estimate reference ET (ET_{ref}), including solar radiation or sky cover, near-surface air temperature, air humidity and wind speed. ET_{ref} is a vital parameter for most remote sensing (RS) approaches to ET estimation, and is generally defined as the ET from a particular hypothetical reference crop. Interest in computing ET_{ref} from gridded weather data has led to the creation of two recent products from NOAA, one estimating ET_{ref} using North American Land Data Assimilation System version 2 (NLDAS-2) data (Hobbins, 2016) and another product forecasting ET_{ref} using the National Digital Forecast Database (NDFD) called Forecasted Reference Evapotranspiration (FRET) (Palmer, Osborne, & Hobbins, 2017). Gridded ET_{ref} is also actively being utilized in the EEFlux remote sensing model (Kilic et al., 2014).

Despite the increasing use of gridded weather data to estimate ET_{ref}, there is a lack of information regarding the accuracy of ET_{ref} estimates that are based on the gridded data, and whether they can successfully be used to replace estimates produced using ground-based weather station data that represent weather in well-watered agricultural settings. Two previous studies that computed ET_{ref} using the ASCE standardized Penman-Monteith equation (ET_{sz}) with data from the NLDAS gridded weather data set found that ET_{sz} was often overestimated by the gridded data when compared to ET_{sz} computed using weather stations in the western continental United States (CONUS) (Lewis et al. 2014; Moorhead et al. 2015). The ET_{sz} overestimation was the result of overestimation of near-surface temperature, solar radiation and wind speed and underestimation of vapor pressure for weather conditions associated with agricultural areas. Studies investigating the weather variables in other gridded data sets found biases like those found in the NLDAS data that should similarly cause overestimation of ET_{sz}. For example, Slater (2016) found that NLDAS, the Global Land Data Assimilation System (GLDAS) and the Climate Forecast System Reanalysis (CFSR) all overestimate incoming shortwave radiation at most locations in the CONUS.

Motivated by the lack of research regarding ET_{ref} estimates made using gridded weather data, the first objective of this thesis was to examine and compare estimates of ET_{ref} made from several gridded data sets to ET_{ref} made from agricultural weather station data across the CONUS. Results of this CONUS-wide analysis should be useful when utilizing ET_{ref} estimated from a gridded data set. Analyses for this objective focused on summer periods because that is when plants are actively growing and ET_{ref} values are at their highest.

The second objective of this study was to compare near-surface air temperature, vapor pressure, shortwave solar radiation and wind speed (the data required for computing ET_{ref}) from gridded data sets to weather station data. This objective is useful not only to professionals estimating ET but to anyone interested in using one of the gridded weather data sets. This objective also helps diagnose the causes of ET_{ref} error. Special emphasis is placed on bias (systematic error) because random error tends to shrink when data are time-averaged.

Unlike the first objective, there have been a variety of studies that have pursued the second objective, comparing gridded weather data parameters to independent weather station data. These studies are discussed in detail in chapter 3. Each study had varying dimensions of scope: extent of study area, study time period, number of weather parameters evaluated, number of gridded data sets evaluated, number of weather stations used, selection of weather stations, and the type of analyses. The differences between previous studies made integrating the information difficult, and provided further motivation to make fresh comparisons.

The gridded data sets analyzed in this research were the Global Land Data Assimilation System (GLDAS), the North American Land Data Assimilation System (NLDAS), the Climate Forecast System version 2 (CFSv2) operational analysis, the Real-Time Mesoscale Analysis (RTMA), the Gridded surface meteorological data set from the University of Idaho (GRIDMET) and the National Digital Forecast Database (NDFD).

This thesis is presented as follows: Chapter 2 – Background, Chapter 3 – Previous Work, Chapter 4 – Methods, Chapter 5 – Results & Discussion, and Chapter 6 –

Conclusions. Chapter 2 provides information about how gridded data are produced, additional information about ET_{ref} and details about each of the gridded weather data sets. Chapter 3 reviews previous research comparing gridded data to weather station data. Chapter 4 discusses how the data were gathered, processed and analyzed. Chapter 5 presents and interprets the results of the data comparisons.

CHAPTER 2. BACKGROUND

2.1 Origins of Gridded Weather Data

Gridded weather data represents estimates of weather parameter values. The values are generally presented as averages within each grid cell, with each grid cell representing a specific area of the earth's surface. Gridded weather data have the advantage of representing weather at regular timesteps and at fixed spatial scales, and over large areas (sometimes globally).

Gridded weather data are produced in a variety of ways depending on their intended applications, but they all generally originate from some combination of observed ground and atmospheric profile data with models. The combination of information from models with observations is called data assimilation (Lahoz, Khattatov, & Menard, 2010). The weights given to the observations and the specific type of atmospheric circulation and land surface models (LSMs) that are employed can vary widely among data sets. For example, the Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 2008) relies heavily on observations that are spatially interpolated using empirical models developed on the basis of relationships between elevation and the estimated weather parameters. Other gridded weather data rely

on atmospheric circulation models and LSMs that are blended with observations. Such data products are typically referred to as analyses.

Meteorological analyses were originally developed to produce initial conditions required by numerical weather prediction (NWP) models. The initial conditions provide what are called "forcings," which is a term referring to any information driving a model that originates from outside of a model (Rood, 2010). The forcings provide a starting point for the simulation models. Though NWP models are forecasting models, the analyses used for the initial conditions also typically employ forecasting models. The analyses may be produced by a making a short forecast using an atmospheric circulation model, which serves as a first guess, and then updating the first guess using observed data (Houser, Lannoy, & Walker, 2010). The short initial forecast that is modified using observation data is called the analysis; a term that refers both to the process and to the product.

The analyses used for NWP (these can be termed operational analyses) have evolved to become extremely useful gridded weather data sets for use in other types of applications, and, along with NWP, are under continual improvement. Unfortunately, however, improvements made over time to assimilation and forecast systems, aimed at producing the most accurate forecasts possible, can create temporal discontinuities in the data. Discontinuities can be problematic for certain models and investigations that look for trends over time. Reanalyses are analyses that are performed using constant assimilation systems, that is, assimilation systems that are not modified over time, to overcome the issues of discontinuity (Saha et al., 2010).

2.2 Evapotranspiration & Reference Evapotranspiration

Evapotranspiration (ET) is a major component of the water cycle, accounting for as much as two thirds of the precipitation falling on land (Wang & Dickinson, 2012). Accurate estimation of ET is crucial to a wide variety of disciplines, but is difficult to directly observe. In addition, ET can vary widely with location due to variation in weather and water availability and variation in vegetation type, health and density. ET has even been shown to vary within a single agricultural field (Irmak et al., 2011). Eddy covariance towers, Bowen ratio systems and lysimeters provide relatively direct estimates of evapotranspiration, but, due to high costs and operational requirements, are sparsely distributed (Long, Longuevergne, & Scanlon, 2014). Remote sensing (RS) and Land Surface Model (LSM) based approaches can solve the problem of spatially sparse ET estimates by estimating spatial variation of ET over large areas. These methods have the potential to provide information for nearly all areas of the globe.

LSMs are models that estimate the transfer of mass, energy and momentum between the earth's surface and the atmosphere (Rodell et al., 2004). Some LSM models are coupled with atmospheric circulation models so that the models can provide feedback to each other. For example, diminishing soil moisture can reduce estimated ET in an LSM and increase estimated sensible heat flux to the air. That process, in turn, causes the atmosphere to warm, which increases the evaporative demand at the surface. Some LSMs are uncoupled from atmospheric models to avoid unrealistic feedback caused by known bias in the atmospheric models. In both cases, LSMs are forced with

meteorological analysis data and therefore depend to a large degree on the accuracy of the analyses used during the production of ET estimates.

Remote sensing approaches to estimating ET generally fall into two categories: surface energy balance approaches that use surface temperature to compute evapotranspiration as a residual of the surface energy balance; and vegetation index based approaches that combine ET_{ref} estimate with crop coefficients estimated from satellite imagery. The vegetation index based approach is essentially a modernized version of the crop coefficient approach pioneered in the FAO24 publication (Doorenbos & Pruitt, 1977). Examples of the surface temperature-surface energy balance approach include the SEBAL (Bastiaanssen, Menenti, Feddes, & Holtslag, 1998), METRIC (Allen, Tasumi, & Trezza, 2007) and EEFlux (Kilic et al., 2014) models. EEFlux is an automated implementation of the METRIC model. ET_{ref} is critical to both types of RS methods for ET estimation. ET_{ref} is also used in more traditional and localized ET estimates for irrigation management.

ET_{ref} is defined in the ASCE-EWRI Task Committee Report (Walter et al., 2005) as "… the ET rate for a uniform surface of dense, actively growing vegetation having specified height and surface resistance, not short of soil water, and representing an expanse of at least 100m of the same or similar vegetation." There is a wide variety of ET_{ref} equations in use. In this thesis we chose to use the ubiquitous ASCE standardized Penman-Monteith equation. We refer to this equation and its estimates in the text as ET_{sz}. Two reference crops are defined for the ET_{sz}, a tall crop similar to 0.5m tall alfalfa and a short crop similar to 0.12m tall cool-season clipped grass (Walter et al., 2005;

Jensen & Allen 2016). In this thesis we computed ET_{sz} using the tall crop which is usually abbreviated as ET_r. The short reference crop ET is usually abbreviated as ET_o.

The equation and details for its correct application are available in Walter et al. (2005). The equation is defined as:

$$ET_{sz} = \frac{0.408\Delta(R_n - G) + \gamma \frac{C_n}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d u_2)} \quad (2.1)$$

Definitions for the parameters are provided in Table 2.1.

Table 2.1: ET_{sz} parameter definitions

Parameter	Definition
ET_{sz}	standardized reference crop evapotranspiration (mm d^{-1})
R_n	calculated net radiation at the crop surface (MJ m^{-2} d^{-1})
G	soil heat flux density at soil surface (MJ m^{-2} d^{-1})
T	mean daily air temperature at 1.5 to 2.5-m (°C)
u_2	mean daily wind speed at 2-m (m s^{-1})
e_s	saturation vapor pressure at 1.5 to 2.5-m (kPa)
e_a	mean actual vapor pressure at 1.5 to 2.5-m (kPa)
Δ	slope of saturation vapor pressure-temperature curve (kPa $°C^{-1}$)
γ	psychrometric constant (kPa $°C^{-1}$)
C_n	numerator constant that changes with reference type and calculation timestep (K mm s^3 Mg^{-1} d^{-1})
C_d	denominator constant that changes with reference type and calculation timestep (s m^{-1})

Computation of ET_{sz} using data collected outside of reference conditions violates certain assumptions underlying the Penman-Monteith equation and can result in an inflated ET_{sz}. Reasonable representation of reference conditions for a weather station are defined in Allen (1996) as a "…well-watered … surface below and surrounding the weather equipment for a distance of 100 times the height of the wind, air temperature and

relative humidity sensors in all directions." Those constraints helped guide our selection of weather data.

2.3 Reasons for Differences Between Gridded and Weather Station Data

This section provides some information about the potential causes of differences between weather station measurements and gridded data estimates. The causes of differences are either due to differences in what the data sets represent or due to error in one or both data sets.

We first discuss differences between the data sets arising from differences in what they represent. Even if gridded data perfectly represent each weather parameter for the time and space they are supposed to represent, there can still be potential problems stemming from using the gridded data to estimate ET_{ref}. One problem arises from the fact that ET_{ref} should be computed using weather data collected over a reference surface. A reference surface is defined by ASCE-EWRI (Walter et al., 2005) as an extensive surface of managed vegetation, such as clipped grass, that is well-watered and actively growing and transpiring. In arid areas with irrigated agriculture, the grid cell size in some gridded data sets might be so large as to extend beyond an irrigated area and include surrounding dry area. The grid cell weather properties would thereby blend the microclimatic conditions of the irrigated area (representing reasonable reference conditions) with non-reference conditions found in the unirrigated environment, potentially diminishing the validity of the ET_{ref} estimate. This is a problem in spatial representivity; the gridded data does not have a small enough spatial resolution to best serve ET_{ref} estimation.

The problem of spatial representivity can also arise in heterogeneous, complex terrain. The gridded weather data sets generally represent the weather conditions at the average elevation of the grid cell. Air properties can change rapidly with elevation, especially air temperature, so if the elevation within the grid cell where ET_{ref} is needed is substantially different from the average grid cell elevation, the data will not represent what is desired.

Besides the problem of spatial representivity, there can be problems associated with temporal representivity. Some gridded data have a temporal resolution (timestep) as large as 6-hours. Given that most gridded weather data sets estimate instantaneous values occurring at each timestep, a 6-hour timestep is analogous to using a 6-hour sampling rate. All four parameters used to estimate ET_{ref} - air temperature, humidity, wind speed and solar radiation - tend to vary substantially during a day, and a 6-hour sampling rate might be insufficient for representing those variations. The coarse temporal resolution could reduce the accuracy of ET_{ref}. Hupet & Vanclooster (2001) found that even an hourly sampling rate could cause errors in daily grass ET_{ref} as large as -27%.

Spatial and temporal representivity impacts can be problematic regardless of the quality of a gridded weather data set, but the quality of a data set itself can also come into question due to inaccuracies in models or in assimilated data. One source of bias expected in gridded data appears because irrigation data is not assimilated. In arid areas, where irrigation is prevalent, the microclimate near the irrigated area is affected by the additional water input. Evapotranspiration made possible by the additional water inputs absorbs solar energy that in the nearby unirrigated areas heats the air instead. The

microclimate near the irrigated crops is therefore cooler and more humid than surrounding unirrigated land, but because the LSMs do not account for water inputs besides precipitation, they overestimate the air temperature and underestimate the humidity.

Another suspected source of bias in in gridded data comes from bias in the assimilated data. Many of the ground-based weather stations assimilated by analyses are located at airports in the CONUS. Temperature and humidity might be elevated at airports because the prevalence of paved surfaces. On pavement there is little to no moisture to vaporize so more solar energy is available to warm the air instead. This is commonly known as the urban heat island effect.

2.4 Gridded Weather Data Sets

This section provides background for each of the gridded weather data sets examined in this thesis. Section 2.1 provides the necessary information to understand this section. The spatial and temporal attributes of the gridded data are shown in Table 2.2.

Table 2.2: Resolutions and coverage of gridded data sets

Data Set	Spatial Resolution (km)	Spatial Resolution (°)	Temporal Resolution (hr)	Temporal Coverage	Spatial Coverage
GLDAS 1	~28 latitude	1/4 (Noah forcings)	3	2000 - present	Global
CFSv2	~22 latitude	~0.2	6	1979 - present	Global
NLDAS 2	~14 latitude	1/8 (Noah forcings)	1	1979 - present	CONUS
GRIDMET	~4.5 latitude	1/24	24	1979 - present	CONUS
RTMA	2.5	-	1	08/22/2006 - present	US, Guam, Puerto-Rico
NDFD 1 day forecast	5 *	-	3 *	12/2/2002 - present	US, Guam, Puerto-Rico
NDFD 7 day forecast	5 *	-	6	12/2/2002 - present	US, Guam, Puerto-Rico

* NDFD spatial resolution changes to 2.5 km and the temporal resolution for forecasts out to 3 days changes to 1 hour after 08/19/2014

2.4.1 GLDAS

Both NLDAS and GLDAS fall under the Land Data Assimilation Systems (LDAS) project of NASA and the National Centers for Environmental Prediction (NCEP) (Mitchell et al., 2000). The goal of the LDAS is to produce LSM's uncoupled from atmospheric models to avoid biases due to coupling. Inadequate coupling can cause feedbacks that create systematic hydrologic error (Houser et al., 2010). Instead the LSM's are forced with high quality data sets derived from observations. These meteorological forcing data sets were utilized in this study, rather than the output of the LSMs.

GLDAS is developed jointly by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Prediction (NCEP) (Rodell et al., 2004). The goal of GLDAS is to generate global high quality fields of land surface states and fluxes using LSMs (Rodell et al., 2004).

GLDAS version 1 primarily derives its forcing fields from the Global Data Assimilation System (GDAS). Precipitation comes from disaggregated Merged Analysis of Precipitation (CMAP), and incoming radiation is derived from the Air Force Weather Agency (AFWA) (Pearce, 2016a). The forcing data include precipitation, downward shortwave radiation, downward longwave radiation, near-surface air temperature, near-surface specific humidity, near-surface zonal wind, near-surface meridional wind and surface pressure (Rodell et al., 2004).

The GDAS is the assimilation system used in support of the Global Forecast System (GFS). The GFS is an operational weather forecast model, which means that its assimilation system (GDAS) is subject to improvements so that the forecasts can be as accurate as possible. This is important because it means that the forcing fields in GLDAS improve over time. The forcing data for GLDAS version 1 drive four different land surface models: Mosaic, Noah, the Community Land Model (CLM) and the Variable Infiltration Capacity model (VIC). In this study we use the forcing data that are used in for the Noah model. Near real-time fields are produced at temporal resolutions of 1-month and 3-hours with a spatial resolution of 1° going back to 1979 and 0.25° (approximately 28 km of latitude) going back to 2000 (Pearce, 2016b).

2.4.2 NLDAS

NLDAS is produced by NOAA/NCEP's Environmental Modeling Center, Climate Prediction Center, NOAA/NWS Office of Hydrological Development, NASA's GFSC, Princeton University and the University of Washington. NLDAS, like GLDAS,

seeks to produce high quality LSM data sets. Similar to GLDAS, NLDAS produces forcing data by assimilating high quality observation based data.

NLDAS version 2 derives most of its forcing fields from the North American Retrospective Reanalysis gridded data set (NARR; Mesinger et al., 2006). To create NLDAS, NARR is spatially and temporally disaggregated employing the methods presented in Cosgrove et al. (2003). Precipitation is derived from a gauge-only climate prediction center (CPC) analysis of daily precipitation orographically adjusted using the PRISM climatology. The PRISM data set is produced by using methods which were briefly outlined in Section 2.1.

One major difference between NLDAS and GLDAS is the primary source of data. For NLDAS, the primary source is a real-time continuation of NARR called the Regional Climate Data Assimilation System (R-CDAS), which is a 'frozen' assimilation system. NARR finished production in 2003, at which time it was transitioned into the R-CDAS (Mesinger et al., 2006). None of the subsequent improvements to data assimilation have been incorporated into the R-CDAS, but they have been incorporated into the GDAS.

NLDAS version 2 produces two different sets of forcing data. In this study we focused on the traditional forcing fields in forcing file A. The forcing parameters produced in forcing file A include precipitation, downward shortwave radiation, downward longwave radiation, near-surface air temperature, near-surface specific humidity, near-surface zonal and meridional wind, surface pressure, fraction of total precipitation that is convective, convective available potential energy and potential evapotranspiration.

NLDAS runs the same four LSM's that GLDAS runs. The forcings and products have a temporal resolution of 1-hour and a spatial resolution of 0.125° (approximately 14 km of latitude) extending back to 1979.

2.4.3 GRIDMET

GRIDMET is produced at the University of Idaho by combining the spatial attributes of PRISM data with temporal attributes of NLDAS version 2 (Abatzoglou, 2011). In GRIDMET, NLDAS is spatially downscaled from its native 14-km grid to a 4.5-km grid and is upscaled from an hourly to a daily timestep. Monthly PRISM data are then used to bias correct the NLDAS data. Only humidity, temperature and precipitation are bias corrected.

The climate data used to debias NLDAS are the PRISM 30-year monthly normals from 1981-2010. In essence, the debiasing procedure aggregates NLDAS data to monthly values and compares the results with PRISM monthly values to create scaling factors that are then applied to the raw daily NLDAS data. During our study period, PRISM was ingesting data from several of the weather station networks used in this study. This may have led to spuriously good performance in the comparisons because the data being verified were determined, in part, by the data they were being verified against. However, it should be noted that in GRIDMET, only monthly PRISM climate normals were used to bias correct NLDAS. In fact, GRIDMET was validated against some of the PRISM input data in Abatzoglou (2011), GRIDMET's founding paper.

2.4.4 CFSv2

CFSv2 was developed by NCEP to replace the CFSv1 (Saha et al., 2014). The purpose of CFSv2 is to produce global seasonal and subseasonal climate forecasts for the globe. The development of CFSv2 began with the creation of the Climate Forecast System Reanalysis (CFSR; Saha et al., 2010). CFSR was produced using state-of-the-art forecast and data assimilation techniques. CFSv2 used the CFSR analysis as initial conditions for retrospective forecasts, forecasts made from points of time in the past where the forecasted time has since passed. The retrospective forecasts were used to calibrate the model for real time forecasts. The CFSR was continued on as the real-time operational Climate Data Assimilation System (CDASv2) in order to continue to provide initial conditions for the CFSv2 forecasts.

The assimilation system of CFSv2 is 'frozen', much like the R-CDAS behind NLDAS. This might seem odd because assimilation systems serving forecasts are usually updated on occasion to improve their accuracy, but because CFSv2 is a climate model, temporal consistency is paramount, and updating the system would likely create inconsistency. The CFSR completed production in 2009 so the models and assimilation techniques of that time are the ones still being used.

In this study the data examined come from the Climate Data Assimilation System version 2 (CDASv2), which is often referred to as the CFSv2 operational analysis. We use CFSv2 to refer to the analysis. The spatial resolution of this data set is 0.2 degrees (approximately 22 km of latitude) and the temporal resolution is six hours.

The Noah LSM (also run by NLDAS and GLDAS) is used in CDASv2 in two ways. First it is used within a fully-coupled land and atmosphere model to make first-guess forecasts. Second it is used in a semi-coupled CFSv2 implementation of GLDAS to perform a land surface analysis. The CFSv2-GLDAS is forced with CFSv2 atmospheric data assimilation output and several combined precipitation analysis data sets. The use of observed precipitation is what makes the model only semi-coupled, and is intended to reduce the bias in CFSv2 precipitation and the resulting unrealistic feedback to and from the Noah model. The CFSv2-GLDAS analysis only occurs once a day at 0000 Coordinated Universal Time (UTC) because the precipitation data are only available at a daily timestep.

Exploratory plots of some CDASv2 parameters versus observations show several distinct trend lines. An example is given in Figure 2.1 of air temperature, where air temperature data from a station in California (designated CA4) are compared with CFSv2 data at the CFSv2 6-hour timestep.

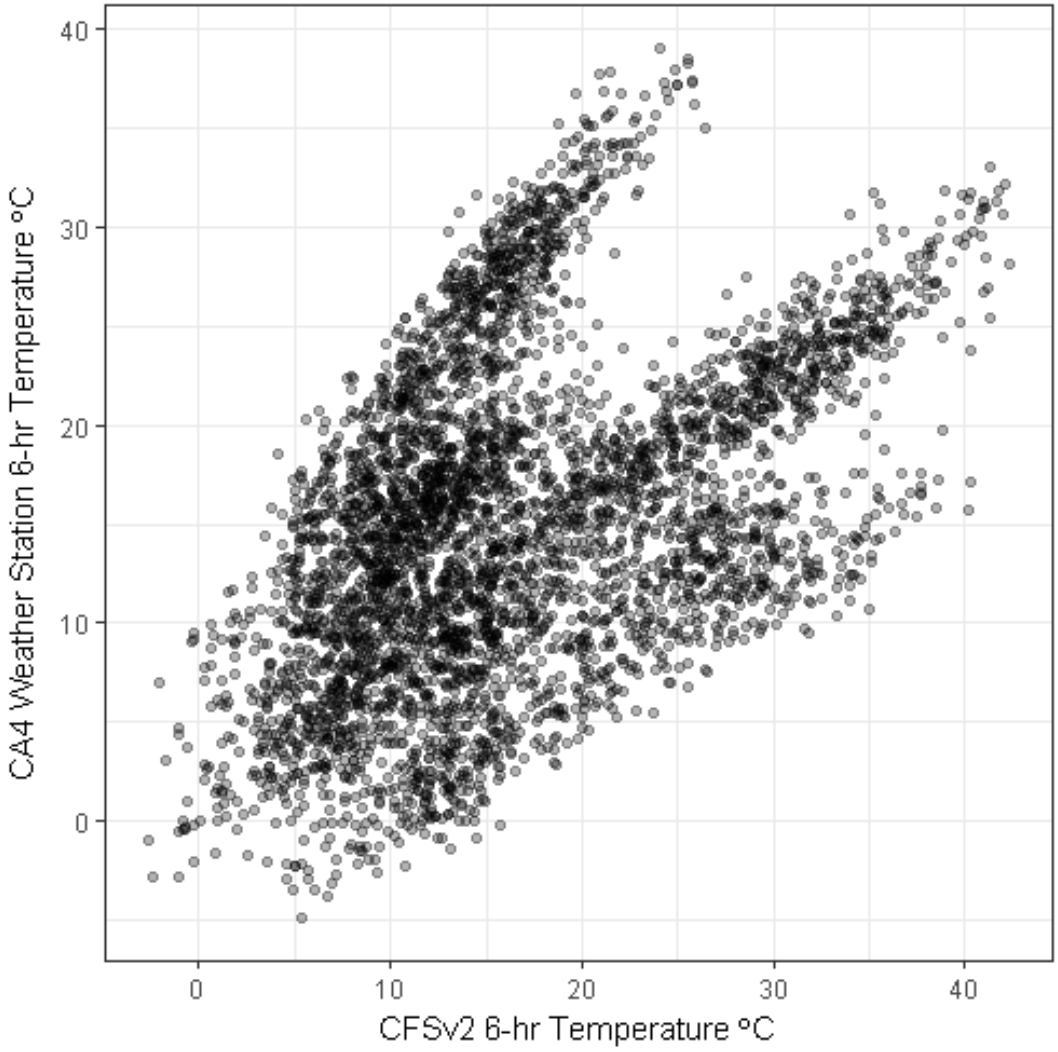

Figure 2.1 6-hr timestep temperature data from 2013 through 2015 from a California weather station plotted against CFSv2 data.

The plot contains data from the entire study period (roughly three years). We hypothesize that the pattern is related to the once-a-day analysis of GLDAS. After the GLDAS analysis for the day, there are three more timesteps for the day where no observed precipitation is assimilated. The analyses at timesteps after the GLDAS analysis rely only on atmospheric model simulated precipitation. If the model simulated precipitation does not correspond well to the GLDAS precipitation analysis (which is closer to reality) then the LSM might diverge from reality during the model simulated timesteps. In this thesis the data are aggregated to daily values. The multiple trends for CFSv2 disappear at a daily timestep, which confirms that the effect is diurnal.

2.4.5 NDFD

NDFD (Glahn & Ruth, 2003) is a product of the National Weather Service (NWS) that began being distributed in 2003 as a means to provide public access to weather forecasts from a central digital source. Weather Forecast Offices (WFOs) produce regional forecasted grids of sensible weather elements that are then mosaicked together on a central server. Forecasters have the ability to manually adjust the forecast grids produced by the numerical forecast (Myrick & Horel, 2006). The nominal spatial resolution of the gridded elements is 5 km and the temporal resolution is 3 hours for short range forecasts with lead times less than 3 days, and 6 hours for longer range forecasts with lead times of 3 to 7 days (Glahn & Ruth, 2003). In late August of 2015 the NDFD reduced its spatial resolution to 2.5 km and its short range forecast temporal resolution to 1 hour. The forecast is produced for the CONUS, Puerto Rico, Hawaii, Guam and Alaska. RTMA, described in the next section, is often used to initialize and verify NDFD forecasts (De Pondeca et al., 2011).

2.4.6 RTMA

The Real-Time Mesoscale Analysis (RTMA) is produced by NCEP, the Earth System Research Laboratory (ESRL) and the National Environmental, Satellite, and Data Information Service (NESDIS) (De Pondeca et al., 2011). The most prominent application of RTMA is to help initialize and verify NDFD forecasts. RTMA has a spatial resolution of roughly 2.5 km and the temporal resolution is one hour, matching those for NDFD.

RTMA is under continual improvement, and periodically assimilates new data sources. RTMA assimilates data from a vast number of ground-based weather stations including many of the networks providing ground-based data used in this thesis. As indicated for GRIDMET, using weather stations to verify a gridded product created using those same weather stations, can cause overestimation of accuracy due to lack of complete independency. Despite this, stations in networks that are assimilated in RTMA were used for our study due to their regional representativeness, data quality and access.

The lack of independence of data sets may not be as problematic as it appears at first glance. First, it is acknowledged that during creation of analyses the modeled weather values are not simply replaced with observed values, but rather a statistical algorithm decides how the two information sources are to be blended. The result is that comparison of assimilated stations to the gridded product following their assimilation can still help to identify biases that exist between the two data sets. Secondly, RTMA ingests many stations, for example, a single temperature analysis for 1500 UTC 20 November 2009 used 14,299 stations (De Pondeca et al., 2011). This means that the analysis would

probably change very little if the RTMA assimilated weather stations used as ground-truth in this thesis were not assimilated in RTMA.

CHAPTER 3. PREVIOUS WORK

The goal of this chapter is to describe findings of previous studies comparing gridded data sets and parameters used in this thesis to ground based observations. Studies outside of CONUS are not reviewed because the quality of gridded products can vary widely by location (Slater, 2016), and the results are not expected to represent conditions for our study area. Some studies included other gridded data sets in addition to the ones focused on in this thesis and used methods that make comparison of results difficult. In this chapter the focus is placed on findings for weather parameters explored in this thesis, and on results that provide grounds for comparison. Bias is quantified in most studies including this thesis and unlike other performance statistics has a sign. The convention that applies to all biases reported in this thesis is that a positive bias means that the gridded data have values that are higher than those of the weather station data.

3.1 GLDAS

Decker et al. (2012) evaluated a number of reanalysis data sets, including CFSR and GLDAS, by comparing them against data from 33 FLUXNET towers located in North America. Comparisons were made for shortwave radiation, temperature, wind speed, precipitation, net surface radiation, latent heat flux and sensible heat flux. FLUXNET is a global network of flux towers that measure weather parameters at specific locations using eddy covariance methods. The systems are designed to measure fluxes of carbon dioxide and energy exchange between earth's surface and the atmosphere. The 33 towers selected were located in a variety of different climate

regimes. The timespan of the data used varied among flux towers, but all comparisons were made between 1996 and 2006. GLDAS version 1 data with 1° spatial resolution were evaluated.

The statistics produced included bias, standard deviation of the error, Pearson correlation, ratio of the standard deviations and root mean squared error (RMSE). The 'ratio of the standard deviations' is an expression of error that describes differing variability about the mean for the two data sources compared. Decker et al. (2012) observed that MSE (and therefore RSME) can be decomposed to a function of correlation, mean of the observed data, mean of the gridded data, standard deviation of the observed data and standard deviation of the gridded data. They used the terms from the decomposed MSE to compute relative contributions to the error for different timescales.

The statistics were computed for each FLUXNET station for three different timescales of the data produced with different aggregation methods: 6-hour data, which was the native timestep for most data sets; the average of each 6-hour timestep for each month (monthly mean diurnal cycle); and monthly averages.

The Decker study presented histograms showing the number of stations falling into numerical bins for bias, RMSE, correlation and ratio of standard deviations, and generally focused on creating rankings for the reanalysis data sets. No precise numerical values for bias or other performance statistics were reported for comparison.

Decker et al. (2012) concluded that at the 6-hour timestep, GLDAS overestimated temperature variability more often than GLDAS underestimated it. The ratios of standard

deviations fell within histogram bins ranging from 0.7 to 1.3, with most stations falling in the 0.9 to 1.1 bin. Almost all correlations fell within the histogram bins ranging from 0.9 to 1.0. Most bias fell within bins ranging from -2.5 to 2.5 °C with one station higher than that range. There did not seem to be a generally positive or negative air temperature bias in GLDAS.

For wind speed, the paper only presents the histogram for the monthly average wind speed. They state that the GLDAS 6-hour data yielded similar results to the monthly data, but that the correlations were lower for the 6-hour data and the spread in the variability histogram was wider. Variability of the monthly data was better estimated by GLDAS than all other gridded data sets. The ratios of standard deviations fell within histogram bins ranging from 0.5 to 5.0 with the relative majority of stations falling within the 0.7 to 0.9 bin. The greatest number of correlations were in the bins ranging from 0.4 to 0.7. Some correlations were negative. The wind speed biases were generally positive and fell between histogram bins ranging from -2.5 to 2.5 m s^{-1}, with most stations in the 0.5 to 1.5 m s^{-1} range.

For solar radiation, the paper only presented the histogram for the monthly average shortwave solar radiation, so the information is less detailed for the 6-hour timesteps. GLDAS ratios of standard deviations were most often between 0.9 and 1.1 but greater than 1.1 at five locations. Most correlations fell between 0.90 and 0.95. The 6-hour shortwave radiation in GLDAS had a small positive bias that we can assume was similar to the monthly bias. The monthly bias fell within histogram bins ranging from -25 W m^{-2} to 50 W m^{-2}.

The comparison of the contributions to MSE of the different components on different timescales showed that at the monthly timescale, bias was the main contributor to error for solar radiation, air temperature and wind speed. The tendency for bias to dominate was stronger in the case of wind speed. For wind speed at the 6-hour timestep, bias often remained a major source of error, but random error played a larger part as indicated by a lower correlation. Decker et al. (2012) results will be commented on again in the CFSv2 Section 3.4, and differences between CFSv2 results and GLDAS results will be discussed.

Slater (2016) compared shortwave radiation from numerous reanalyses, satellite and derived products against 4000 weather stations within the CONUS and parts of southern Canada. NLDAS version 2, GLDAS version 1, and CFSR were all evaluated. The gridded products were compared against the observations using the bias and correlation of the radiation on a daily timestep. Data from between the years 2000 and 2010 were used, though the exact period of data available at each station varied.

Slater calculated what he referred to as a clear sky ratio (CSR) in order to correct and rescue biased weather station solar radiation measurements. CSR was defined as the ratio of weather station radiation on a typical clear sky day to the theoretical clear sky radiation value as calculated using the Bird and Hulstrom (1981) model. CSR was calculated for overlapping periods of time at each weather station and the value closest to 1.0 for a particular period was used as an adjustment factor.

Slater found that the radiation bias during the summer months (June, July and August) were positive for NLDAS and GLDAS except for low or negative bias in the

Rockies and Northwest. Maps of bias showed that bias ranged from -45 W m^{-2} to 45 W m^{-2}. GLDAS reached a positive bias as high as the 35 to 45 W m^{-2} range, but the negative bias only reached the -25 to -35 W m^{-2} range. These biases matched up well to the biases found in Decker et al. (2012).

Finally Slater assessed variability in the gridded solar radiation by correlating the clear sky index (measured radiation divided by theoretical clear sky radiation on all days, including clouded days) of the measured data against the clear sky index of the gridded product data. The clear sky index was used to indicate the ability of the gridded products to model impacts of cloudiness. GLDAS had a correlation between 0.70 and 0.75. These correlations are generally lower than the 6-hour correlations found in Decker et al. (2012), but the sub-daily timestep in Decker et al. should increase the correlation relative to a daily timestep because the sub-daily timestep will track the diurnal cycle. Furthermore, correlating the clear sky index should result in a lower correlation than using the radiation data.

3.2 NLDAS

There have been more studies exploring error in NLDAS than for all other gridded data sets. In Slater (2016), NLDAS had a clear sky index correlation between 0.65 and 0.70. The summertime radiation bias was positive for NLDAS except for the Rockies and the Northwest, a pattern shared with GLDAS. Also, like GLDAS, NLDAS reached a positive bias as high as the 35 to 45 W m^{-2} range, but the negative bias only reached the -25 to -35 W m^{-2} range.

Lewis et al. (2014) validated NLDAS version 2 to specifically assess the viability of using NLDAS data for computing ET_{sz}. The study only evaluated locations in the western United States and for weather stations located in agricultural environments, where ET_{sz} is typically applied. They gathered data from 704 stations, from 1979 to 2012. The parameters evaluated were shortwave radiation, air temperature, wind speed, relative humidity and ET_{sz}. During their data processing, Lewis et al. used spatial bilinear interpolation of NLDAS pixel values to the weather station locations to reduce error due to differences in the spatial scale. They also adjusted temperature and humidity to the weather station elevations. Their analysis was performed at the hourly timestep of NLDAS.

The statistics computed to validate the NLDAS data included the coefficient of determination (R^2), bias and RSME. Air temperature and solar radiation had the highest R^2 values ranging from 0.76 to 0.96 and 0.77 to 0.93 respectively, across locations. R^2 for relative humidity ranged from 0.27 to 0.72, and for wind speed ranged from 0.14 to 0.61. R^2 for air temperature tended to be lower in southern Texas, the Rockies, and the Pacific coast. R^2 for solar radiation was relatively low in Montana, the Pacific Northwest, Southern Texas and the northern Great Plains. For wind speed, the lowest R^2 values occurred in the Rockies, the Southwest and the Pacific Northwest. Relative humidity presented a spatial pattern similar to the wind speed, except that the central valley in California and most of Arizona had high R^2.

Solar radiation bias ranged from -7.5 W m^{-2} to 33.6 W m^{-2}, near-surface air temperature from 0.1 °C to 4.0 °C, wind speed at 2 m height from –1.5 m s^{-1} to 1.7 m s^{-1}

and near-surface relative humidity from -20.8% to 2.5%. Solar radiation bias was highest from California down through Arizona and southern New Mexico, with additional high points in central Oregon and in Montana and North Dakota along the Canadian border. The radiation biases computed by Slater showed a slightly different spatial pattern, perhaps due to restricting the computation of bias to the summer months. The main differences in bias occurred through the eastern parts of Nebraska, Kansas, Oklahoma and Texas; Slater et al. showed those areas as having a high bias, contrary to Lewis et al. who showed a negative bias.

It is worth emphasizing that near-surface air temperature bias was positive everywhere in the Lewis study. The highest biases occurred in southern California, Nevada, Arizona and the Rockies. Wind speed biases were highly variable spatially, with negative biases in the Great Plains north of Kansas and with higher biases in eastern Kansas, Oklahoma and Texas, and additionally on the West Coast, extending from Northern California to Canada. The spatial patterns in relative humidity bias closely matched the patterns for air temperature.

The bias for grass reference ET_{sz} (ET_o) was positive for most of the country and ranged between -0.007 mm h^{-1} and 0.052 mm h^{-1}. California and the areas along the Mexican border had the highest positive biases. The R^2 ranged from 0.81 to 0.93. Interestingly the R^2 was highest along the border with Mexico reflecting precision in timing of trends in ET_o, but bias in estimation. R^2 was lowest in Montana, the Pacific Northwest, southeastern Texas, Nebraska, South Dakota and North Dakota.

A study closely related to Lewis et al. (2014) is Moorhead et al. (2015), in which the ET_{sz} maps produced by NOAA using NLDAS data were evaluated in the Texas High Plains region. They compared ET_{sz} values, minimum temperature, maximum temperature, solar radiation and wind speed between NLDAS/NOAA and 14 Texas High Plains ET network weather stations. The comparisons were made using hourly data for the time period of 2001-2010. Linear models fitted between the data sets, RMSE and the Nash-Sutcliffe Efficiency (NSE) were used in the comparisons.

At all weather stations in the Moorhead study, the linear model for ET_{sz}, where ET_{sz} from gridded data were regressed against ET_{sz} computed from weather stations, produced slopes and intercepts showing that the use of NLDAS data generally overestimated ET_{sz} (a slope of 1 and an intercept of 0 would show that NLDAS was unbiased). The slopes and intercepts found for daily minimum and maximum air temperature showed that they were typically overestimated by NLDAS. This result agrees with Lewis et al. (2014). The results from Moorhead et al. (2015) and Lewis et al. (2014) also agree on wind speed bias, where there was a negative bias in the Texas panhandle. The regression coefficients found for solar radiation did not show a pattern, and no generalizations could be made. Generally the results confirmed the findings of the earlier study by Lewis et al.

3.3 GRIDMET

GRIDMET is a relatively recently created gridded weather data set, so evaluations of GRIDMET appear infrequently in the literature. The principal reference for GRIDMET is Abatzoglou (2011), which both described how GRIDMET was produced

and how validation was applied for some of the products of GRIDMET over the western United States (approximately west of the eastern New Mexico border).

The Abatzoglou study compared GRIDMET data to data from 1618 weather stations from several different networks including AgriMet. Some of the ground stations used in this thesis were part of the AgriMet network. Abatzoglou compared maximum and minimum temperature and relative humidity (RH), precipitation, wind speed, vapor pressure deficit, energy release component (ERC) and ET_o computed with the standardized Penman-Monteith method. The performance statistics were computed using daily data, and included the Pearson correlation coefficient, mean absolute error (MAE) and mean bias. Each statistic was computed for the cool season (Oct-Apr) and warm season (May-Sept). GRIDMET temperature and humidity were not lapse-adjusted for station elevation within a GRIDMET grid cell, and this was noted as an additional source of error in the paper.

Daily maximum and minimum near-surface air temperatures had median correlations of 0.94-0.95 and 0.87-0.90, and MAE of 1.7-2.3 °C. Bias maps of minimum and maximum daily temperature show that bias for both temperatures ranged from -2 to 2 °C. The Abatzoglou paper stated that stations from the Remote Automated Weather Stations (RAWS) weather station network produced a median minimum temperature bias of -0.95 °C in contrast to non-RAWS stations which had a bias of -0.12 °C. They hypothesized that the reason the air temperatures in the gridded weather data set tended to be cooler than measured air temperature at RAWS stations is due to the tendency of RAWS stations to be situated in areas having fire risk that often coincide with areas in the

thermal belt. Thermal belts appear in mountainous regions where, at night, a band of warm air becomes trapped mid-slope by cold-air drainage. Another interesting observation made by Abatzoglou was that many stations that were located less than 10km from each other often had very different minimum temperature correlations when the surrounding terrain was heterogeneous. He attributed the difference in correlations to the inability of gridded data to reasonably account for cold air drainage, perhaps due to the terrain homogenization within each grid cell.

Minimum and maximum RH had median correlations between 0.77 and 0.81 among the weather stations and median MAE between 6 and 12%. In general, daily maximum RH had lower correlations than daily minimum RH. The spatial patterns in the error were similar to those for air temperature.

Wind speed had median correlation values of 0.54 during the cool season and 0.52 for the warm season. Abatzoglou found that wind speed tended to have a high bias in forested areas. He noted that some studies have found topography and surface roughness to cause substantial differences between observed and interpolated wind.

ET_o in Abatzoglou (2011) had a median correlation among weather station locations of 0.90 but showed a median bias of positive 0.5 mm d^{-1}. He attributed the high ET_o bias to biases in the other variables. Abatzoglou also noted the possibility of bias appearing especially high at AgriMet weather stations due to the irrigation effects mentioned in Chapter 2. A time series plot of ET_o averaged across 7 AgriMet stations and ET_o computed using GRIDMET shows a substantial bias averaging 14% high for the period of March to October pooled across the study years.

Another study comparing GRIDMET data to weather station data is McEvoy et al. (2014). The aim of that study was to assess the ability of gridded data to represent weather along large elevation gradients. Weather station data were taken from transects in the Snake and Sheep mountain ranges of Nevada from October to September of 2012. The variables analyzed included daily minimum and maximum temperature, daily average dew point temperature and daily total precipitation. The evaluation statistics used included bias, R^2 and MAE.

In the Snake range, daily maximum air temperature biases were positive at higher elevations and negative in the valleys which they attributed to differences between the grid pixel elevations and the station elevations - one of the spatial representivity problems addressed in Section 2.3. An interesting result with daily minimum temperature was that stations located in the foothills were warmer than the valley floor stations due to the cold air drainage and thermal belt mentioned by Abatzoglou (2011). GRIDMET was able to reproduce the nighttime air temperature inversions that occur in valley areas, but underestimated the strengths of the effect. Dew point bias wasn't observably affected by elevation and tended to be negative in both the cold and warm seasons with a magnitude of no more than 3 °C. The results from the Sheep range were similar. For R^2, maximum and minimum temperature and dew point all performed worse during the cold season (Oct-Mar) than during the warm season (Apr-Sept).

3.4 CFSv2

Slater (2016), previously mentioned in reference to GLDAS and NLDAS comparisons, also evaluated the solar radiation product from the Climate Forecast System

Reanalysis (CFSR). The CFSR data tend to be very close to the operational CFSv2 analysis data set, so conclusions reached for CFSR are also applicable to the CFSv2 analysis. Slater found that CFSR-based solar radiation had a positive bias throughout most of CONUS for all seasons, with exceptions in small pockets of West Texas and the Pacific Northwest. The bias spanned the range of -45 W m^{-2} to 45 W m^{-2}.

Decker et al. (2012), previously mentioned in reference to GLDAS, also evaluated CFSR. The CFSR temperature ratios of standard deviations fell in histogram bins ranging from 0.7 to 1.3, with most stations falling in the 0.9 to 1.1 bin. CFSR underestimated variability at more stations than did GLDAS. The correlations were greater than 0.90 for most stations for all gridded products. Like GLDAS, CFSR overestimated 6-hour near-surface air temperature roughly as frequently as it underestimated it. The bias fell in histogram bins ranging from -3.5 to 6.5 °C.

A 6-hour wind speed histogram was not presented, but they reported that results were similar to the monthly wind speed. Monthly wind speed variability was usually overestimated by CFSR. The ratios of standard deviations fell in histogram bins ranging from 0.5 to 5, with most stations in bins greater than 0.9. The greatest number of correlations were in the bin from 0.4 to 0.7. Some correlations were negative, and a couple were above 0.90. Bias in wind speed was positive for CFSR, as it was for GLDAS. Bias fell into histogram bins ranging from -0.5 to 3.5 m s^{-1}

A 6-hour solar radiation histogram was not presented, so the information is less detailed. The ratios of standard deviations were between 0.9 and 1.1 at most stations. Correlation was reported to be between 0.9 and 0.95 at most stations. CFSR had a

positive solar radiation bias, similar to GLDAS. The bias fell in histogram bins ranging from -25 to 50 W m^{-2}.

Decker et al. (2012) gave each gridded weather product a ranking for bias and the standard deviation of the error at each of 33 FLUXNET stations. Since there were six gridded products the rankings were 1-6. The rankings were then averaged across all stations to provide a consolidated comparison among the gridded products.

The rankings showed GLDAS to be superior to CFSR in terms of bias for temperature and wind speed, but worse than CFSR for solar radiation. The rankings for standard deviation of error showed GLDAS to be superior to CFSR for wind speed only, with air temperature and solar radiation being better in CFSR.

3.5 NDFD

Myrick & Horel (2006) computed a variety of forecast quality assessments on NDFD temperature data from the winter of 2003-2004 in the western CONUS. The goal of their study was to decide what metrics would be useful for verifying NDFD forecasts, to examine differences between verification against observations and verification against analysis data sets, and to estimate forecast improvement as a function of lead time. The weather station data used for verification came from the MesoWest data repository of the University of Utah. The spatial extent of the comparisons was the eleven states west of Nebraska. Two verifying analysis data sets were used. One was the Advanced Regional Prediction System Data Assimilation System (ADAS) created at the University of Utah (Xue et al. 2000, 2001, 2003). The other analysis was the Rapid Update Cycle (RUC) provided by NCEP (Benjamin et al., 2004).

Myrick & Horel used many different approaches to explore the validity of the NDFD forecasts including some methods used in this study. Bias and RMSE were computed between NDFD and the verification analysis data sets, and between NDFD and the observation data for different forecast lead times for forecasts issued at 0000 UTC. The performance statistics were very similar for all the verification data sets. Bias did not greatly increase as a function of forecast lead time, but RMSE did. Bias oscillated between being positive and negative for every 12 hour increment of lead time. A 12-hour lead time typically had a bias around -1.5 °C for near-surface air temperature, with bias computed using the weather station data. A 24 hour lead time had a bias around 0.6 °C (a 36 hour lead time returned to a bias of -1.5 °C). The RMSE increased with forecast lead time, with even multiples of a 12-hour lead time, having relatively smaller RMSE but still increased with lead time.

The authors theorized that the oscillations in bias were due to the method employed by forecasters to estimate near-surface air temperature; they first estimate daily minimum and maximum air temperature and then interpolate to find all other temperature values. The western states are 7 or 8 hour behind UTC, which places the 12 hour forecast lead time near the daily minimum temperature and the 24 hour forecast lead time near the daily maximum temperature. The authors stated that minimum and maximum temperatures persist only briefly. If the interpolation overestimates the duration of the minimum and maximum temperatures that could result in the underestimation of air temperature in the morning and overestimation in the afternoon shown in the study.

It should be emphasized that the results of Myrick & Horel (2006) represent only one winter season in the western United States, and more importantly, that NWS forecasts (and therefore NDFD) have been improving since the time period analyzed by Myrick and Horel (winter of 2003-2004). Furthermore, the temperature field is not used in our study, but rather minimum and maximum temperature are averaged to estimate a daily average temperature.

Another paper comparing NDFD forecasts, Perez et al. (2010), focused on solar radiation at seven sites in CONUS. As will be further discussed in the methods chapter, NDFD does not contain shortwave solar radiation forecasts; instead it contains sky cover estimates. In Perez et al. (2010) they estimated solar radiation from sky cover using an equation developed in Perez et al. (2007). The empirical equation from Perez et al. (2007) was refitted to data from the seven evaluation sites to improve the equation's performance beyond a 3-day forecast lead time to which the equation was originally fit. They reported that there was a tendency for sky-cover under-prediction beyond the 3-day forecast. Although NDFD has a coarser temporal resolution, sky cover was time interpolated to create hourly estimates of solar radiation.

The comparison of ground station data came from the SURFRAD network operated by the NOAA Earth System Research Laboratory (ESRL). The seven sites were located in Desert Rock, Nevada; Fort Peck, Montana; Boulder, Colorado; Sioux Falls, South Dakota; Bondville, Illinois; Goodwin Creek, Mississippi; and Penn State, Pennsylvania. The study used data from August 2008 to August 2009. Performance

statistics included bias, RMSE and the Kolmogorov-Smirnov Integral (KSI) goodness of fit test.

Perez et al. (2010) analyzed the daily RMSE and bias at lead times of one to seven days for different time groupings: all year, winter, spring, summer and fall. The comparisons were made on an hourly timestep. Bias in daily solar radiation was generally negative for all stations except the Montana station which only had a negative bias in winter. Bias was also negative for all seasons except fall; only the Mississippi station had negative bias in the fall. There was generally an increase in the magnitude of the bias from 1 to 7 day lead times though there were station and seasonal exceptions. Averaging the bias across all seasons and stations yielded a bias of -14 W m^{-2} at a 1-day lead time and -22 W m^{-2} at a 7-day lead time.

The RMSE was lowest in Nevada, which the author attributed to there being fewer clouds to predict in that climate. RMSE also increased from the 1-day to 7-day lead times. Averaging the RMSE across all seasons and stations yielded an RMSE of 149 W m^{-2} for a 1-day lead time and an RMSE of 191 W m^{-2} for a 7-day lead time.

3.6 RTMA

Only one study was found that compared RTMA against surface observations (Charney, Zhong, Kiefer, & Zhu, 2013). That study remains unpublished because the verifying ground data were taken from the RAWS network; a network assimilated into RTMA. It is difficult to find independent weather observations to compare with RTMA because RTMA assimilates data from many weather networks. The goal of the Charney et al. (2013) study was to compare meteorological parameters important in wild fire

management between RAWS, a network frequently used for fire management, and RTMA in order to better understand how to use RTMA and NDFD in fire management and forecasting. The study used data from August 2008 to July 2010 collected at 237 RAWS stations in the northeastern CONUS. They computed a wide variety of statistics but did not report most of them.

The mean bias was positive for wind speed, and was negative for relative humidity. Charney et al. (2013) found that the near-surface air temperature bias varied between stations, but that it was usually plus or minus 2 °F (1.1 °C). Mean bias for wind speed fell between -1 and -3 mph (-0.5 and -1.3 m s^{-1}) for most stations, and for relative humidity it fell between 2 and 4%. They reported no significant seasonal or spatial patterns in the bias.

CHAPTER 4. METHODS

The general flow of the data is shown in Figure 4.1. Gridded and weather station data were processed to meet the two objectives of this thesis. Each objective required slightly different processing, which is reflected in Figure 4.1 by the two separate terminations to the flowchart. The details of the steps shown in Figure 4.1 are explained in the body of this chapter.

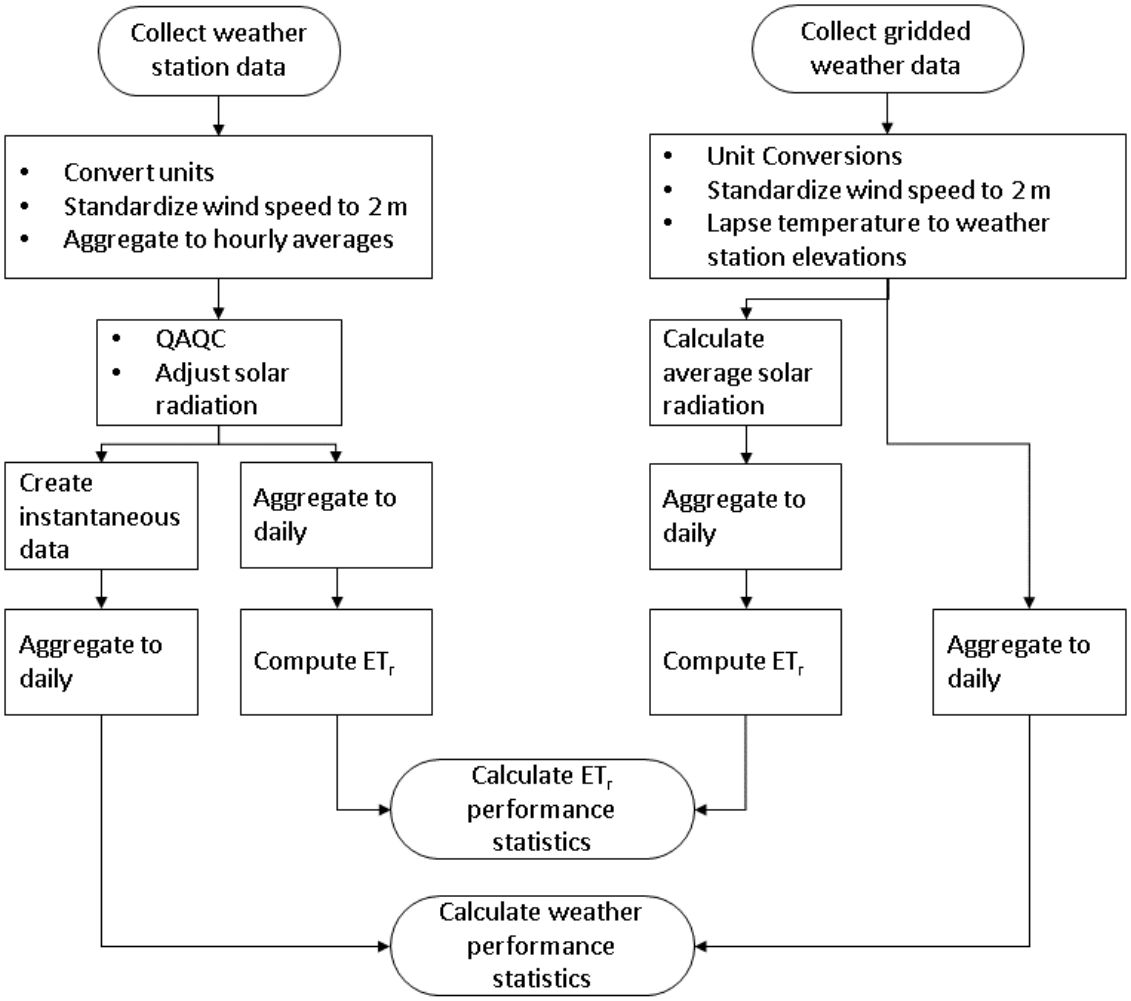

Figure 4.1 Data processing flow chart.

4.1 Station Selection

The selection of the 103 ground weather stations, used for comparisons to gridded data, was guided by a number of considerations. Efforts were made to select weather stations representing reference conditions where possible because those are the conditions under which ET_{sz} estimates are the most valid and error may be unique for those conditions. Reference conditions tend to occur in arid regions where there is irrigated agriculture, which is where ET_{sz} estimates are often applied for irrigation water

management. In addition, in remote sensing applications such as METRIC and EEFlux, it is important to use ET_{sz} estimates that originate from well-watered reference settings due to the impact of ET_{sz} on the calibration and accuracy of those models. Section 2.3 described how dry and warm biases in gridded data are expected in irrigated areas due to the lack of assimilated irrigation data.

Selected weather stations were required to measure all parameters evaluated (temperature, humidity, shortwave solar radiation, and wind speed) at a maximum timestep of an hourly rate for the evaluation period of 2013 through 2015. Furthermore, station networks with free, open data access were preferred to networks requiring payment. Lastly, stations were selected that represented reference or near-reference siting conditions where possible. Reference conditions, as previously described, are the conditions under which the ET_{sz} equation is valid. In the real world these conditions can be difficult to find or verify. To help identify these conditions, the Normalized Difference Vegetation Index (NDVI) was used. NDVI is a standard vegetation index that is a commonly derived product from satellite imagery. High values for NDVI indicate the presence of live green vegetation, which generally indicates the occurrence of high soil moisture, potential ET rates and associated cooling of the air. NDVI is calculated as shown in equation 2.1

$$NDVI = \frac{(\rho_{NIR} - \rho_{VIS})}{(\rho_{NIR} + \rho_{VIS})} \quad (4.1)$$

where ρ is reflectance, NIR stands for the near infrared band and VIS is the visible red band. Band widths for NIR and VIS can vary among instruments, so NDVI can vary from satellite instrument to satellite instrument. Plants strongly absorb visible red light

but strongly reflect light in the near infrared spectrum, so an NDVI of 1.0 indicates very lush vegetation completely covering the entire pixel. NDVI less than around 0.2 often corresponds to nearly bare soil. NDVI is usually negative for pixels containing clouds, water or snow.

Although NDVI can be high for vegetation suffering water stress, low NDVI indicates a lack of green, actively transpiring vegetation. In other words, a high NDVI is a necessary but not completely sufficient indication of reference conditions. Despite that limitation, NDVI was used as the primary tool for identifying stations that were likely to be near reference conditions.

The approach to selecting weather stations was iterative. First, average NDVI for the warm season months of April through September was calculated for each study year from MODIS MCD43A4 16-day surface reflectance composites using the Google Earth Engine (Google Earth Engine Team, 2015). MODIS stands for Moderate Resolution Imaging Spectroradiometer and is an instrument aboard the NASA Terra and Aqua satellites. Pixel sizes for that product are 500 m by 500 m. Negative pixels were masked out. Monthly values were extracted from the 500 m pixels over the station locations. These single pixel values at the station locations are referred to as local NDVI. An average of all 500 m pixels within 2 km of the stations, termed areal NDVI, was also computed. The means and standard deviations of all monthly NDVI's (point and areal) were then calculated. Stations having a mean point and areal NDVI's less than 0.4 were discarded. For stations with coordinates rounded to the hundredths place, only the areal NDVI was used as a criteria because of the uncertainty of the actual station location.

After removing stations having low NDVI, the best stations, in terms of NDVI, in the geographic regions desired were selected. This procedure also involved some judgement: weighing between NDVI, the location, and the variability in the NDVI.

Once candidate stations were selected, satellite images from Google Earth and ground based images from Google Maps (https://www.google.com/maps) or taken by the station network, if available, were consulted to judge the station surroundings. Each station was evaluated for reference surroundings, i.e. few obstructions and uniform low vegetation extending in all directions for approximately 100 times the height of the temperature and humidity sensors. Stations not meeting those requirements were removed from the list of candidates to produce the list of final candidates. If the removal of a station due to insufficient surroundings caused a large region of CONUS to go unrepresented, another station was selected from among the available stations even if the alternative did not exhibit reference conditions. Doubling back to replace stations constituted the iterative portion of the process. Poor quality data from a station were also grounds for choosing an alternative station. The final set of weather stations contained some stations that are closely surrounded by trees and others that have low NDVI, because there were no alternative weather stations in those regions that had better represented reference conditions.

Figures 4.2 and 4.3 show the mean NDVI and standard deviation of the monthly NDVI for the study period for both the point and areal NDVI. Areal NDVI values were generally lower in the west because of the incorporation of dry land within the 2 km radius.

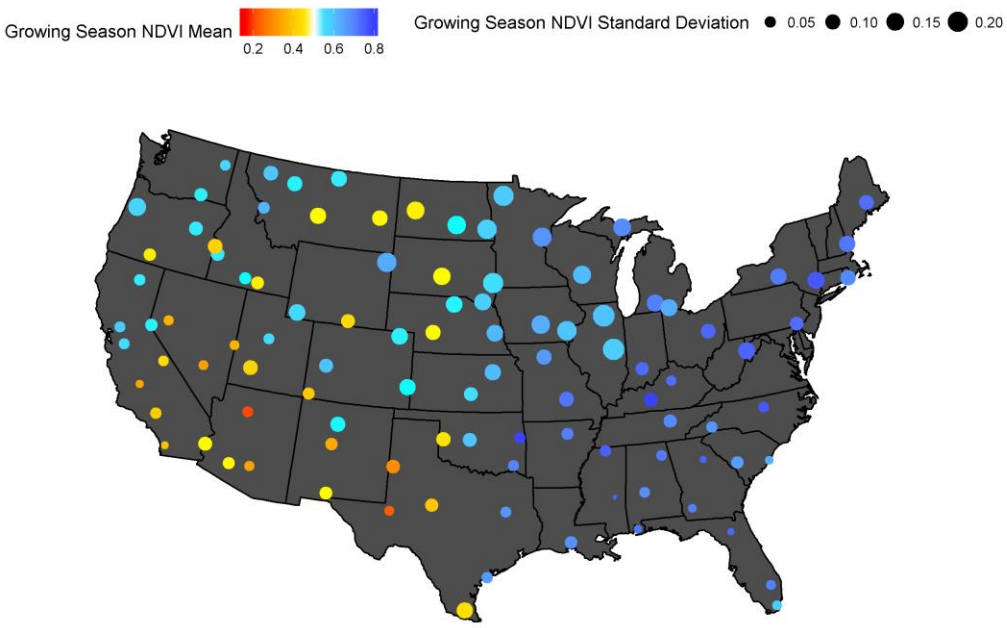

Figure 4.2 Means and standard deviations of monthly NDVI within a 2 km radius of weather stations during the warm season (Apr-Sep) pooled across study years.

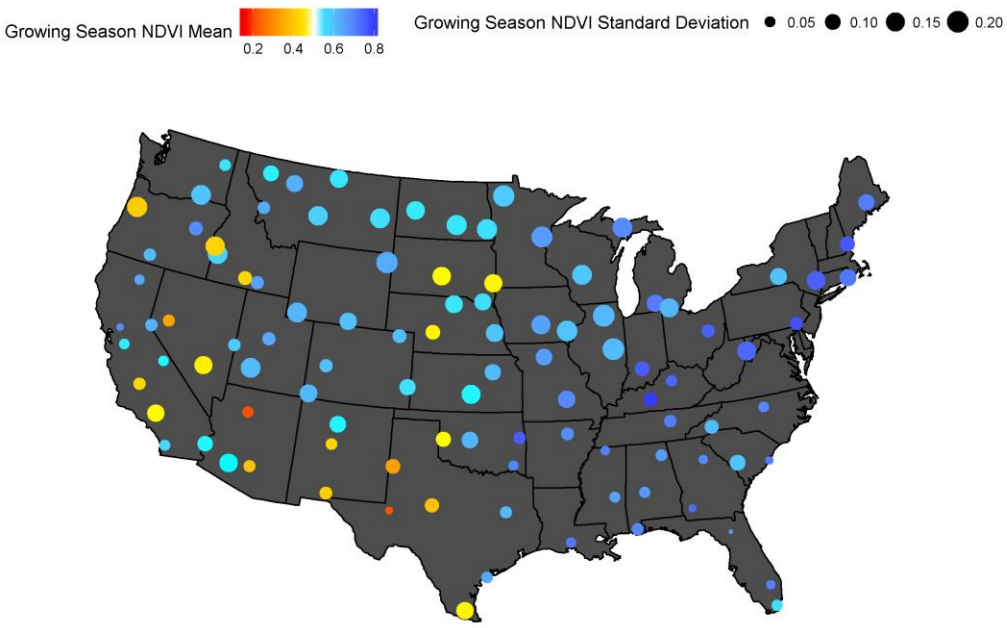

Figure 4.3 Means and standard deviations of monthly point NDVI at station locations during the warm season (Apr-Sep) pooled across study years.

4.2 Data Retrieval & Extraction

4.2.1 Weather Station Data

Weather station data were downloaded from fifteen different sources that represented thirteen different weather station networks. Some networks were state-wide for single states, for example the FAWN network for Florida, and others were regional, for example the AgriMet Cooperative Agricultural Weather Network operated by several entities. Some networks required email requests to obtain data but most provided an online data access service. A total of 103 stations were selected. These stations were located in 41 US states and in all US climate regions.

Data were typically in comma or whitespace delimited text files. Python scripts were developed to parse data from each source and to manipulate the data into a standard format. Basic metadata were retrieved for each station including the heights of the anemometer, the heights of the temperature and humidity instruments, and the station latitude, longitude and elevation. Table 4.1 shows the networks and data sources accessed.

Table 4.1: Sources of weather station data

Weather Networks	Network Abbreviations	Data Sources	Number of Stations Used
United States Climate Reference Network	USCRN	USCRN	47
Arizona Meteorological Network	AZMET	AZMET	2
California Irrigation Management Information System	CIMIS	CIMIS	8
Florida Automated Weather Network	FAWN	FAWN	3
Colorado Agricultural Meteorological Network	CoAgMET	High Plains Regional Climate Center	4
Iowa Environmental Mesonet	IAM	High Plains Regional Climate Center	1
Kansas State University	Kansas State	High Plains Regional Climate Center	1
Enviro-weather Automated Weather Station Network	Enviro-weather	Enviro-weather	2
North Dakota Agricultural Weather Network	NDAWN	NDAWN	3
Nebraska Mesonet	NEmesonet	High Plains Regional Climate Center	6
New Mexico Climate Center	NMCC	NMCC	2
AgriMet Cooperative Agricultural Weather Network	AgriMet	AgriMet Great Plains, AgriMet Pacific Northwest, Desert Research Institute, Utah AgWeather Network	20
Oklahoma Mesonet	OKMesonet	OKMesonet	4

4.2.2 GLDAS

GLDAS data were downloaded from the Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/). The data are stored in GRIB files (GRIB stands for GRIded Binary) which is a standardized file format devised by the World Meteorological Organization (WMO) for storing weather data. GRIB files were downloaded for the Noah model with 0.25° spatial resolution and a 3-hour time step. Only the Noah model files contain parameters at a spatial resolution of 0.25 degrees. Each file contains raster grids for all the forcing fields and LSM fields valid for one timestep.

Once all files for the study period were downloaded to a local directory, the values of pixels overlying the station locations were programmatically extracted. Data for all parameters at each station location were written to a comma separated value (csv) file, with one csv per station location.

4.2.3 NLDAS

The NLDAS data were also gathered from the Goddard Earth Sciences Data and Information Services Center. GRIB files of forcing data set A (described in section 2.4.2) were programmatically downloaded with 0.125° spatial resolution and a 1-hour timestep. Like GLDAS, each file contains data for all fields for one timestep. The data were extracted at the weather station locations using the same methods described for GLDAS.

4.2.4 GRIDMET

GRIDMET data were downloaded from the Google Earth Engine, which serves as a repository for GRIDMET data. The ImageCollection ID was IDAHO_EPSCOR/GRIDMET. Data were accessed February 7th, 2017. A script, similar to the one used to extract the NDVI values used for station selection, was created to extract the values of GRIDMET pixels at station locations for our study period.

4.2.5 CFSv2

CFSv2 data were also extracted from the Google Earth Engine. Earth Engine ingests a number of surface meteorological parameters from the NOAA CFSv2 operational analysis 6-hourly surface and radiative flux GRIB files. The ImageCollection ID was NOAA/CFSV2/FOR6H. Data were accessed February 7th, 2017.

The files ingested by Earth Engine are not analysis files, but forecast files; the analysis files contain only eight parameters and are missing several of the parameters investigated in this study so they were not used. The forecast files ingested in Earth Engine show what has been estimated by the model before being updated with observational data. Each forecast file was produced by forecasting 6 hours ahead of the previous analysis.

4.2.6 NDFD

NDFD data span the entire study period but have significant gaps, especially in 2015. The historical NDFD data were gathered from the NOAA National Operational Model Archive and Distribution System (NOMADS) (https://www.ncdc.noaa.gov/data-access/model-data). NOMADS provides an ftp server to download archived NDFD

forecasts. The data are stored in GRIB files. For NDFD, each GRIB file represents a forecast from either one to three or four to seven days for one weather parameter. The files contain bands of data; each band is a raster representing the weather parameter for a single forecast "valid time". NDFD forecasts are published roughly every hour. Therefore, the last forecast and second-to last-forecast of each day were downloaded programmatically from the ftp server in order to obtain the most accurate forecast for the following days. These forecasts were the most accurate because they were initialized as near as possible to the beginning of our forecasted days of interest. The second-to-last of the day forecast was only downloaded for parameters whose 3-day lead time forecasts use 3-hr timesteps, because the last forecast of the day does not forecast the first hour of the following day.

Forecast accuracy decreases with the lead time of the forecast. A homogenous forecast data set therefore requires not only that forecasts represent the same time in the future but that the forecasts are made for the same lead time. The data were filtered to meet these requirements.

The files were filtered according to metadata stored within the GRIB files. For files forecasting 1-3 day lead times the criteria for filtering were that the first forecast should be for 0000 UTC on day 1, and the first forecast duration should be one hour, that is, the forecasted values were predicted from one hour prior. For files forecasting the 4-7 day lead times, the criteria for filtering were that the first forecast should be for 0600 UTC on day four, and the first forecast duration should be three days and six hours. Only files meeting these criteria were used in order to preserve temporal consistency.

Once the all the files for the study period were filtered, the values of pixels overlying the station locations were extracted. Data for all parameters at each station location were combined and written to a csv file, i.e. one csv file per station location and forecast lead time (1-day and 7-day).

4.2.7 RTMA

RTMA data do not cover the entire study period of 2013 through 2015. The data including a sky cover variable (equivalent to what is present in NDFD) began to be produced in mid-April of 2015, so that date is used as the start of the RTMA data analyzed in this thesis.

RTMA data came from two sources. One source was an NCEP ftp server that provides recently produced RTMA data (ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/rtma/prod/). RTMA GRIB files were downloaded from this server by staff at UNL in near-real-time before they disappeared from the NCEP server. The other data source was Google Earth Engine, which only has historical RTMA data back to June 18 of 2015. The ImageCollection ID was NOAA/NWS/RTMA. The download date was February 7, 2017. Data produced before June 18, 2015 are from the GRIB files, data after that date come from Earth Engine. The values of pixels overlying the weather station locations in both data sets were extracted. Afterward, the two time series were joined.

4.3 Data Processing

In order to meet the objective of comparing gridded weather data to weather station data, data processing was required to manipulate the time-averaged weather

station data to match the temporal representation of the gridded data, which usually represented instantaneous values. The manipulations and resampling of data were necessary so that consistent and fair comparisons could be made. Data processing was also needed to meet the second objective of evaluating the accuracy of gridded weather ET_{sz} estimates, because the raw gridded data are not in a form necessary for input into the ET_{sz} equation.

The weather variables compared were vapor pressure, near-surface air temperature, shortwave solar radiation and wind speed. All data were converted to units required by the ET_{sz} equation (see Table 2.1). Wind speed data were standardized to a 2 meter height. An equation assuming a logarithmic wind profile was employed to perform the wind conversions. That equation was taken from Walter et al. (2005).

Both the weather data comparisons and the ET_{sz} evaluations were performed on a daily timestep for all variables. A daily timestep was chosen because aggregating to a day helps to remove bias caused by any temporal misalignments of data. A day is a typical time period over which ET_{sz} is computed, since it is generally the shortest period in which water management decisions are made. Lastly, a day is the smallest common timestep among all data. The data processing performed aggregation to daily timesteps for both the weather data comparison and ET_{sz} evaluation.

Selecting the time zone on which to perform the analysis was challenging because Coordinated Universal Time (UTC) and local time both have advantages. Weather station data are typically recorded on local time, and gridded data, except for GRIDMET, are all reported on UTC. Using UTC allows all data to be in the same time zone,

simplifying some programming tasks. Local time allows ET_{sz} to be calculated as it would be in practice. Ultimately, local time was selected for the analysis.

Aggregation and analysis were performed on local standard time for all gridded data except NDFD and GRIDMET. The smallest GRIDMET timestep available for download is one day and the data are on Central Standard Time (CST) so weather station data were all aggregated on CST to align with GRIDMET. NDFD analysis was performed on Universal Coordinated Time (UTC), and the weather station data with which NDFD were paired were aggregated on the same. The reasons for this are presented in Section 4.3.6.

Minimum and maximum air temperature are required by the daily ET_{sz} equation. Many of the gridded data have maximum and minimum daily temperature determined on UTC. Despite the transition to local time the UTC minimum and maximum temperatures were used because it was judged that, most of the time, UTC minimum and maximum temperature would correspond to the minimum and maximum temperatures found on an arbitrary local time across CONUS.

Pacific Standard Time (PST) is 8 hours behind UTC, Mountain Standard Time (MST) is 7 hours behind UTC, Central Standard Time (CST) is 6 hours behind UTC and Eastern Standard Time (EST) is 5 hours behind UTC. The daily minimum air temperature typically occurs around dawn and the daily maximum air temperature typically occurs in the mid-afternoon. A UTC day completely covers all morning hours across CONUS so that the reported daily minimum temperatures are consistent. The

UTC day ends anywhere from 1600 hours on PST to 1900 hours on EST which should, the majority of the time, include the mid-afternoon high temperature.

For the objective of comparing weather data, before aggregating weather station data to a daily timestep, a linear temporal interpolation was performed to simulate the instantaneous hourly data reported by the gridded data sets. Then the data was temporally resampled to match each gridded data set with which it was paired. This was done so that the weather station data would contain the same amount and type of information as the gridded data. Attempting to match the temporal representivity of the gridded data allowed error to be attributed more directly to the gridded data.

Aggregation of gridded data to a daily timestep for estimating ET_{sz} used instantaneous data, except for solar radiation; no temporal interpolations were made to simulate the average values present in the weather station data. Performing an interpolation of gridded data possessing native timesteps greater than one hour would have resulted in data from neighboring days being blended into the final daily value. This is because the gridded data, when converted to local time, do not evenly subdivide a day (except on CST). The exception is solar radiation, for which interpolations were made. On local time, blending in solar radiation information from a neighboring day isn't a problem because the blending occurs during nighttime when there zero solar radiation.

When evaluating the ET_{sz} based on gridded data, ET_{sz} was compared to ET_{sz} calculated from the time-averaged weather station data, that is, not from the weather data manipulated to match the gridded temporal representation. This is because the aim of the

ET_{sz} evaluation was to determine how well the gridded data perform in estimating the ET_{sz} relative to the best estimates produced from weather stations. This is the reason for the two separate terminations in the data processing flow chart, Figure 4.1.

The differences in spatial representivity between gridded and weather station data have been identified as a problem in the literature and in this thesis. An effort was made to reduce some of these differences by reconciling the vertical spatial representivity. In areas that have complex terrain, the average elevation of a gridded data cell can differ substantially from the elevation of a weather station contained within that grid cell. Temperature generally changes with elevation at a lapse rate of -6.5 °C per 1000 m gained (Cosgrove et al., 2003). We adjusted gridded weather temperatures to the weather station elevations using this lapse rate and the mean elevations of the grid cells. Vapor pressure was converted to relative humidity and then converted back to vapor pressure using the lapsed temperature. Mean grid cell elevations were obtained from the websites for GLDAS, NLDAS and GRIDMET. NDFD, CFSv2 and RTMA elevations were computed by taking the mean of the National Elevation Dataset (NED) within the grid cells.

The last step in data processing was to compute standardized alfalfa ET_{ref} (ET_r). The daily ET_r equation was coded in python and verified by comparing calculations with the REF-ET software of the University of Idaho (Allen, 2016). Processing details specific to each data set are provided in the subsections below.

4.3.1 Weather Stations

The first processing step was to make the weather station data as homogenous among stations as possible. One hour was the smallest common timestep, so all data with smaller timesteps were aggregated to one hour. Precipitation was summed and other parameters were averaged. The desired parameters at this first juncture were temperature [°C], vapor pressure [kPa], solar radiation [W m^{-2}], wind speed [m s^{-1}], hourly maximum temperature [°C], and hourly minimum temperature [°C]. Minimum and maximum air temperature required special treatment as did humidity.

The type of humidity data downloaded followed the preferences in Walter et al. (2005). Vapor pressure, the form of humidity ultimately desired, if not available for download, was computed at the smallest timestep possible so that all subsequent aggregations would be performed using the ideal form for expressing humidity. The form of humidity available varied by weather network.

For some weather networks, separate daily data files were downloaded because they provided the most accurate estimates of daily minimum and maximum air temperature. To get minimum and maximum temperature when the raw data had timesteps smaller than one hour, the min/max temperatures for each hour were computed. Other weather networks included hourly minimum and maximum temperatures. At a later step, when aggregating to a daily value, the daily min/max values were found as the minimum of the hourly minimum temperatures and maximum of the hourly maximum temperatures.

Homogenized, the weather station data were then subjected to quality assurance and quality control (QAQC), where weather station data were plotted for visual QAQC. The visual inspection of the data followed the guidelines suggested in Allen (1996) and Walter et al. (2005). Data that were suspected of being in error were set to a null value.

Solar radiation has the benefit of being extremely predictable during clear sky conditions. This allows data generated from miscalibrated instruments to be relatively evident because they do not fall near the clear sky solar radiation curve even on clear days. The REF-ET software that is distributed by the University of Idaho Kimberly Research and Extension Center comes packaged with a QAQC executable which contains an algorithm to estimate adjustment factors to solar radiation data. We encoded the algorithm in python in order to be able to make custom modifications to it.

The algorithm finds solar radiation measurements that are likely to have occurred during clear-sky conditions and computes the adjustment factor as the mean of the ratio of the measured radiation to the theoretical clear-sky radiation (R_s/R_{so}). Measured radiation is then adjusted by dividing by the adjustment factor.

The algorithm, as we have implemented it, breaks the solar radiation time series into adjustment periods. We matched the REF-ET software and set the adjustment period to 30 days, except for the stations in the Pacific Northwest, which were processed using a 60 day adjustment period to compensate for less frequent clear-sky days. If the last adjustment period is less than the defined length it is combined with the previous adjustment period.

Three other parameters besides the adjustment period length are required during the QAQC of solar radiation and its adjustment. One is the maximum final R_s/R_{so} allowed for any timestep, which we set to 1.2, the REF-ET software default. When R_s/R_{so} exceeds the maximum, the R_s value is set equal to R_{so}. This procedure eliminates high outliers. The second parameter is the number of high values, representing clear sky conditions, to use to calculate the solar radiation adjustment factor. This was set to REF-ET software default of 10 for all stations. The last parameter to set was the proportion of highest R_s/R_{so} values to skip before computing the adjustment factor. This parameter helps account for occurrences high R_s/R_{so} values due to clouds reflecting additional radiation back to the sensor. Using such values could result in adjusted R_s values being too high. Again we used the REF-ET software default value of 0.02 percent. Using those parameters the algorithm averages together the R_s/R_{so} values to determine the adjustment factor.

When an adjustment factor was less than 1% away from 1.0, we deemed that no adjustment was needed, and the adjustment factor was set equal to 1.0. This was done in order to leave data unmodified when possible. The adjusted solar radiation were plotted for visual inspection to ensure that the adjustments were reasonable, and to reverse any poor adjustments. If the radiation data, before adjustment, were *very* far away from the clear sky solar radiation, then other nearby stations were examined for suitability as replacements.

The second-to-last step in the processing of the weather station data was to resample and aggregate the data in order to temporally match the gridded data. During

aggregation, null values were allowed to propagate, i.e. if a single null value was present within an aggregation period, the aggregated value was set to null. This was done to ensure high quality daily values. It was also at this step that wind speeds measured at heights other than 2 meters were converted to equivalent wind speeds at a height of 2 meters, and daily minimum and maximum air temperatures were finally computed. Lastly ET_r was computed for various aggregated versions of the weather station data.

4.3.2 GLDAS

Table 4.2: GLDAS data downloaded

GLDAS			
Parameter	Units	Parameter Representation	Timestamp Location
Air Temperature	°C	instantaneous	-
Specific Humidity	kg kg^{-1}	instantaneous	-
Shortwave Radiation	W m^{-2}	instantaneous	-
U-Component Wind Speed	m s^{-1}	instantaneous	-
V-Component Wind Speed	m s^{-1}	instantaneous	-

When data were processed for creating the ET_r estimates, average 3-hour solar radiation was produced for GLDAS. First, 3-hour average clear-sky radiation for the GLDAS time series was computed. Then instantaneous clear-sky radiation was computed for the time series. Finally, the hourly average clear-sky radiation was multiplied by the ratio of GLDAS radiation to the instantaneous clear-sky radiation. This ratio represents the impact of clouds during the 3-hour period. The morning and evening hours can produce ratios of GLDAS radiation to clear-sky radiation much larger than 1; these ratios were set to 1 before computing the hourly averaged solar radiation.

GLDAS data were then aggregated to daily values on local time. Daily minimum and maximum air temperatures were selected from among the 3-hr instantaneous temperature estimates. Daily time series were produced for the weather data comparison and the ET_r evaluation.

4.3.3 NLDAS

Table 4.3: NLDAS data downloaded

NLDAS			
Parameter	Units	Parameter Representation	Timestamp Location
Air Temperature	°C	instantaneous	-
Specific Humidity	kg kg^{-1}	instantaneous	-
Shortwave Radiation	W m^{-2}	instantaneous	-
U-Component Wind Speed	m s^{-1}	instantaneous	-
V-Component Wind Speed	m s^{-1}	instantaneous	-

Daily minimum and maximum air temperature were selected from among the hourly instantaneous temperature estimates. For calculation of ET_r, the hourly average solar radiation was estimated using the same procedure described for GLDAS in Section 4.3.2. NLDAS data were then aggregated to daily values on local time. Daily time series were produced for the weather data comparison and the ET_r evaluation.

4.3.4 GRIDMET

Table 4.4: GRIDMET data downloaded

GRIDMET			
Parameter	Units	Parameter Representation	Timestamp Location
Daily Minimum Air Temperature	K	-	start
Daily Maximum Air Temperature	K	-	start
Specific Humidity	kg kg^{-1}	average	start
Shortwave Radiation	W m^{-2}	average	start
Wind Speed	m s^{-1}	average	start

GRIDMET processing was relatively simple compared to the other gridded data sets. Since it is a daily data set, no aggregations were necessary. Due to the nature of the data, only one time series was produced for both ET_r evaluation and weather data comparison.

4.3.5 CFSv2

Table 4.5: CFSv2 data downloaded

CFSv2			
Parameter	Units	Parameter Representation	Timestamp Location
Minimum Air Temperature	K	-	start
Maximum Air Temperature	K	-	start
Specific Humidity	kg kg^{-1}	instantaneous	-
Shortwave Radiation	W m^{-2}	average	start
U-Component Wind Speed	m s^{-1}	instantaneous	-
V-Component Wind Speed	m s^{-1}	instantaneous	-

CFSv2 processing was also relatively straightforward. CFSv2 data already contain average solar radiation values so they did not need to be computed. Daily minimum and maximum air temperatures were selected from the 6-hour minimum and maximum temperatures for the day.

4.3.6 NDFD

Table 4.6: NDFD data downloaded

NDFD			
Parameter	Units	Parameter Representation	Timestamp Location
Daily Minimum Air Temperature	°C	-	end
Daily Maximum Air Temperature	°C	-	end
Dew Point	°C	instantaneous	-
Sky Cover	%	instantaneous	-
Wind Speed	m s^{-1}	instantaneous	-

The NDFD data were aggregated to daily values on UTC time. Recall that the data downloaded for NDFD were the last and second to last forecasts made each UTC day. Aggregating the NDFD data on the different CONUS time zones would have mixed data from different NDFD forecasts together, with the effect of having some time zones with more long lead time forecasts than others. Time zones with more values from longer forecasts would produce larger apparent error. Therefore, it was necessary to aggregate NDFD only on UTC.

NDFD, despite having timestamps on UTC, determines daily minimum and maximum temperature on local time. But as previously discussed, daily minimum and maximum temperature should usually be the same whether found on local time or UTC time, so the NDFD minimum and maximum temperatures were used.

For each day of NDFD data, the values at the start of the day (timestamped 0000) and at the very end of the day (timestamped 0000 on the next day) were included in the aggregated daily averages, but those values were given half weight because they were on the edges of the averaging bin. This was a valid approach because NDFD timestamps evenly subdivide a UTC day.

NDFD does not contain shortwave solar radiation values but it does contain sky cover. Two of equations were found in the literature that map sky cover to solar radiation. Those equations were used to estimate average radiation for NDFD so that ET_r could be calculated. The quality of the derived solar radiation data was also evaluated.

The NDFD FRET uses an equation developed from the FAO24 publication by Doorenbos and Pruitt (1977) to compute solar radiation from sky cover (C. Palmer, personal communication, November 9, 2016). The equation is below:

$$R_s = \left(0.25 + \left(0.50 * \left((-0.0083 * Sky\ Cover) + 0.9659\right)\right)\right) * R_a$$

Where R_s is the estimated solar radiation, and R_a is the computed extra-terrestrial solar radiation. We compared the results using the Doorenbos & Pruitt equation to the results using an empirical equation developed in Perez et al. (2007) specifically for NDFD. That equation is:

$$R_s = R_a * \left(1 - 0.87 * ((Sky\ Cover/100)^{1.9})\right)$$

Sky cover was temporally linearly interpolated to simulate the average sky cover for each timestep before being used in the equations. The Perez et al. equation is expected to be worse for the 7-day forecast than for the 3-day forecast because the equation was fit to 1-3 day forecasts only (Perez et al., 2010).

One detail discovered late in our research was that NDFD changed to a finer spatial scale (from 5km to the RTMA grid of 2.5km) on August 19, 2014. At that same time, for forecasts out to the 3-day lead time, wind speed, dew point and sky cover become available at a 1-hour timestep. Ideally, the finer NDFD should be analyzed separately from the coarser resolution NDFD, but to simplify the study, the 1-hour data were sampled at a 3-hour increment in order to match the older data.

2.3.7 RTMA

Table 4.7: RTMA data downloaded

RTMA			
Parameter	Units	Parameter Representation	Timestamp Location
Air Temperature	°C	instantaneous	-
Dew Point	°C	instantaneous	-
Total Cloud Cover	%	instantaneous	-
U-Component Wind Speed	m s^{-1}	instantaneous	-
V-Component Wind Speed	m s^{-1}	instantaneous	-

RTMA was processed in a manner similar to NLDAS, the other hourly instantaneous data set, except that RTMA contains a total cloud cover (sky cover) variable and no solar radiation. RTMA solar radiation, like NDFD radiation, was estimated using Doorenbos & Pruitt (1977) and Perez et al. (2007).

4.4 Performance Statistics

The mean squared error and its variations (e.g. Nash-Sutcliffe efficiency) are commonly used to evaluate the performance of hydrological models by comparison to observed data. Gutpa, Kling, Yilmaz and Martinez (2009) presented a decomposition of MSE into a function of variance, bias, and correlation. A more elegant combination of those components is presented as the Kling-Gupta efficiency (KGE). The model perfectly matches the observations when KGE is equal to 1. KGE is defined in the following equations:

$$KGE = 1 - ED \tag{4.2}$$

$$ED = \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{4.3}$$

$$\alpha = \sigma_M/\sigma_O \tag{4.4}$$

$$\beta = \mu_M/\mu_O \tag{4.5}$$

$$r = \frac{\sum y^O y^M - \frac{\sum y^O \sum y^M}{n}}{\sqrt{\left(\sum y^{O^2} - \frac{(\sum y^O)^2}{n}\right)\left(\sum y^{M^2} - \frac{(\sum y^M)^2}{n}\right)}} \tag{4.6}$$

The superscript M refers to modeled data (gridded data) and the superscript O refers to the weather station (observed) data. ED stands for the Euclidean distance from the ideal point in the scaled space (see Gupta et al. 2009 for details), α is the ratio of the modeled data standard deviation to the observed data standard deviation, β is the ratio of the modeled data mean to the observed data mean and r is the Pearson correlation coefficient.

We also computed more traditional model performance statistics. The statistics computed included mean bias error (MBE), standard deviation of error (SDE) and the correlation (which is also in the KGE and is defined in equation 2.5). The statistics are defined in the following equations.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i^M - y_i^O) \tag{2.7}$$

$$SDE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((y_i^M - y_i^O) - MBE\right)^2} \tag{2.8}$$

Where n is the total number of observations and y_i is the i-th observation. These statistics were computed for the weather data comparisons and for the ET_r evaluations for each gridded data set.

The statistics were not only computed for all available data but were computed for a number of additional groupings: month by year; month pooled across years;

meteorological seasons pooled across years with spring defined as March, April, May, summer as June, July, August, fall as September, October, November and winter as December, January, February; meteorological seasons by year; the warm season, defined as April through September, pooled across years; and the warm season by year. The different groupings of data provide insight about how the error changes seasonally. Groups were required to have at least 50% of their complete data in order to ensure the statistics were computed with enough data to reasonably represent the time period. The numerous groupings resulted in a tremendous amount of results to sift through and necessitated focusing on particular groupings for presentation.

CHAPTER 5. RESULTS & DISCUSSION

5.1 Weather Parameter Comparisons

This results sections focuses on the summer season because summer is the primary growing season for most of CONUS and is the period when ET rates are largest. Results are primarily explored in this section via plots and maps. Plots and maps either show statistics computed for the summer season pooled across all study years or statistics computed for summer 2015 only, so that RTMA, which does not have summer data for 2013 or 2014, can be compared to the other data sets. Maps of bias mark station from networks that are assimilated into the gridded data sets with a box symbol instead of a circle. Boxplots are in the style of Tukey, which shows the median as the center line, the first quartile as the bottom of the box, the third quartile as the top of the box, and the whiskers as extending to the last observation within 1.5 times the interquartile range. The points on boxplots above and below the whiskers are outliers.

Though other model performance criteria (error statistics) were calculated, KGE is most prominent in the discussion because it can be conveniently broken into components, each of which show a different characteristic of the error. Each of those components is equally weighted in KGE which allowed us to make easily interpretable rankings of the gridded data sets.

Depending on the intended use of the gridded data, different error statistics may be more important or useful than others. If one needs an average of ET_r over the entire growing season, then bias (β in KGE) might be most important to minimize because it reflects systematic error. If one is using NDFD to schedule irrigations for the next week, the correlation (r in KGE) and variability error (α in KGE) would be important to consider because they reflect how much error one might expect day to day. Correlation is indicative of random error. Variability error is a systematic error that can be considered similar to the slope of a linear regression line; a slope of 1 means that the gridded and observed data have the same standard deviation.

5.1.1 Near-surface Air Temperature

The temperature bias was positive for all gridded data sets for the bulk of the weather stations, confirming what was found in previous studies. Median biases ranged from positive 0.54 °C for GRIDMET to positive 2.3 °C for NLDAS. Median correlations ranged from 0.87 for GRIMDET (or 0.65 for NDFD day 7 if considering forecasts) to 0.97 for RTMA. One point to note about temperature is that, unlike other weather variables, 0 is not a lower boundary to the possible values temperature can have. This creates an issue for the use KGE β because β cannot be interpreted as a percent bias. The

possibility of negative temperature β creates ambiguity about which mean is actually greater. Fortunately, within CONUS, summertime mean temperatures are typically above 0 °C.

The bias tended to be more strongly positive in the western states (see Figures 5.6 through 5.12). There are a variety of explanations for this spatial trend, many of which were mentioned in the background and previous work chapters but are reiterated here. The spatial heterogeneity of land use, soil water availability and topography is greater in the western US than in the eastern US, which can produce error in multiple ways. Gridded data don't take irrigation into account, which can cause gridded temperatures to be higher than real temperatures. In a related manner, air masses in western states tend to be drier than in eastern states due to lower rainfall, so the evaporative demand and therefore potential for cooling of an air mass is greater than it is in the eastern states when the air mass is in contact with a relatively wet surface. Stations in valleys can experience cold air drainage thereby lowering the average daily temperature, an effect that is difficult for gridded data to replicate (McEvoy et al. 2014).

GRIDMET median bias was smallest among the analysis data sets, excluding RTMA. GRIDMET also had the smallest variability error. The improved performance over NLDAS, one of GRIDMET's parent products, is clear for each KGE component except correlation. The median correlation for GRIDMET temperature was slightly below the correlation for NLDAS. The lower correlation may be related to the debiasing procedure applied to NLDAS to produce GRIDMET. In some cases bias and variability error may have been traded for correlation.

Referring to Figure 5.3, GLDAS and GRIDMET dominated the overall KGE rankings across the country, with CFSv2 taking the lead at a hand full of stations. The good performance by GLDAS can be attributed to its use of the operational GDAS data. Figure 5.5 shows the stability of the KGE performance across the study years. The relative performance of each data set seems constant except for GRIDMET, which performed relatively poorly in 2013. This anomaly allowed GLDAS to take the lead for that year.

When we restrict our view to only 2015 data, so that RTMA can be included in the comparison, we find that RTMA and GRIDMET dominated the top KGE ranking map for temperature (Figure 5.4). We also notice that GRIDMET correlation was better than NLDAS correlation in Figure 5.2 suggesting that the low GRIDMET KGE in 2013 might be primarily due to a low correlation that year. The success of RTMA was not surprising given that the assimilation system is continually improved and assimilates a greater amount of weather station data than the other data sets.

NDFD forecasts performed well relative to the analysis data sets. The 1-day lead time forecasts had generally better performance than NLDAS in all three KGE components. One of the main differences in the pattern of error for NDFD relative to the analysis data sets was that the forecasts tended to underestimate variability while the analysis data sets typically overestimated it. There was a modest increase in bias from the 1-day to 7-day lead times but the increase in error was mostly correlation and variability error reflecting greater uncertainty about the more distant future.

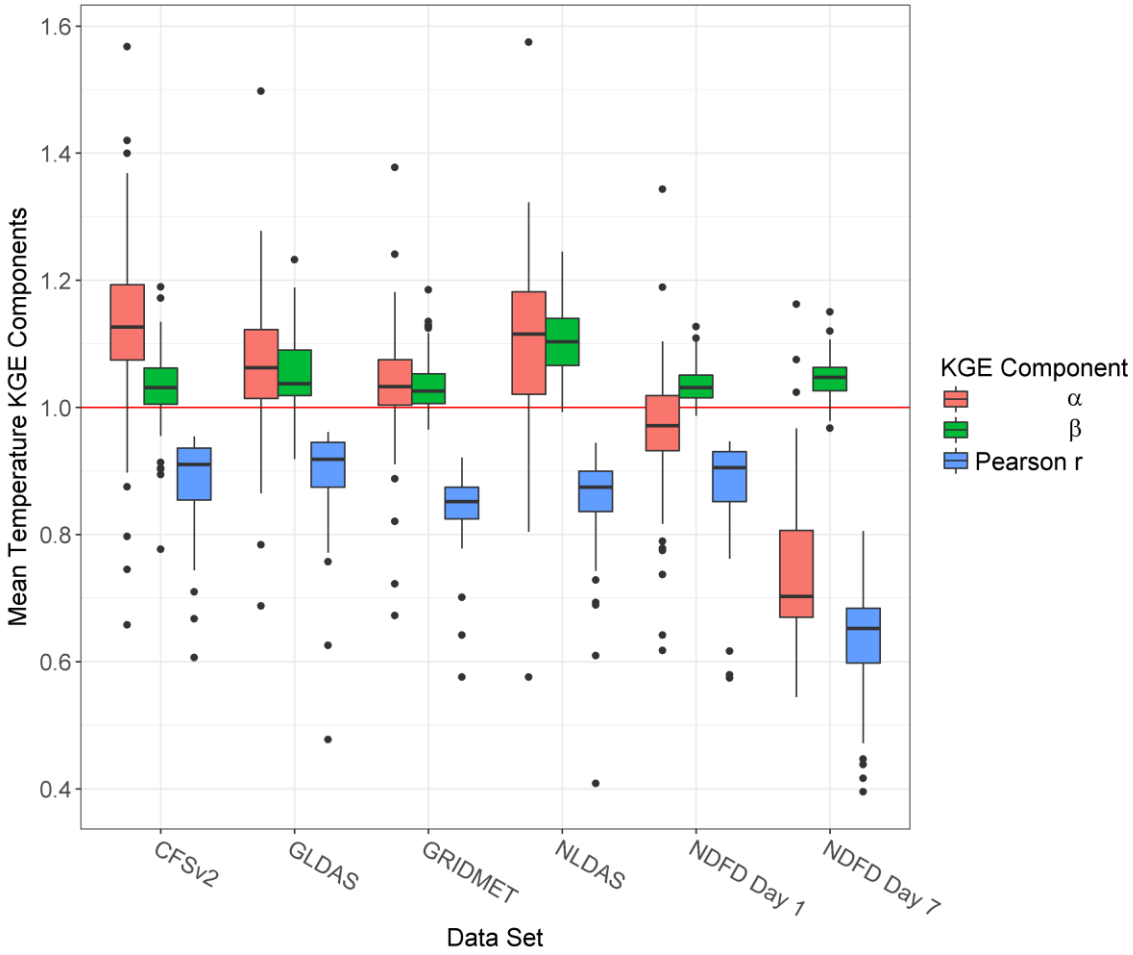

Figure 5.1 Near-surface air temperature KGE components computed using summertime daily data at each station pooled across all study years.

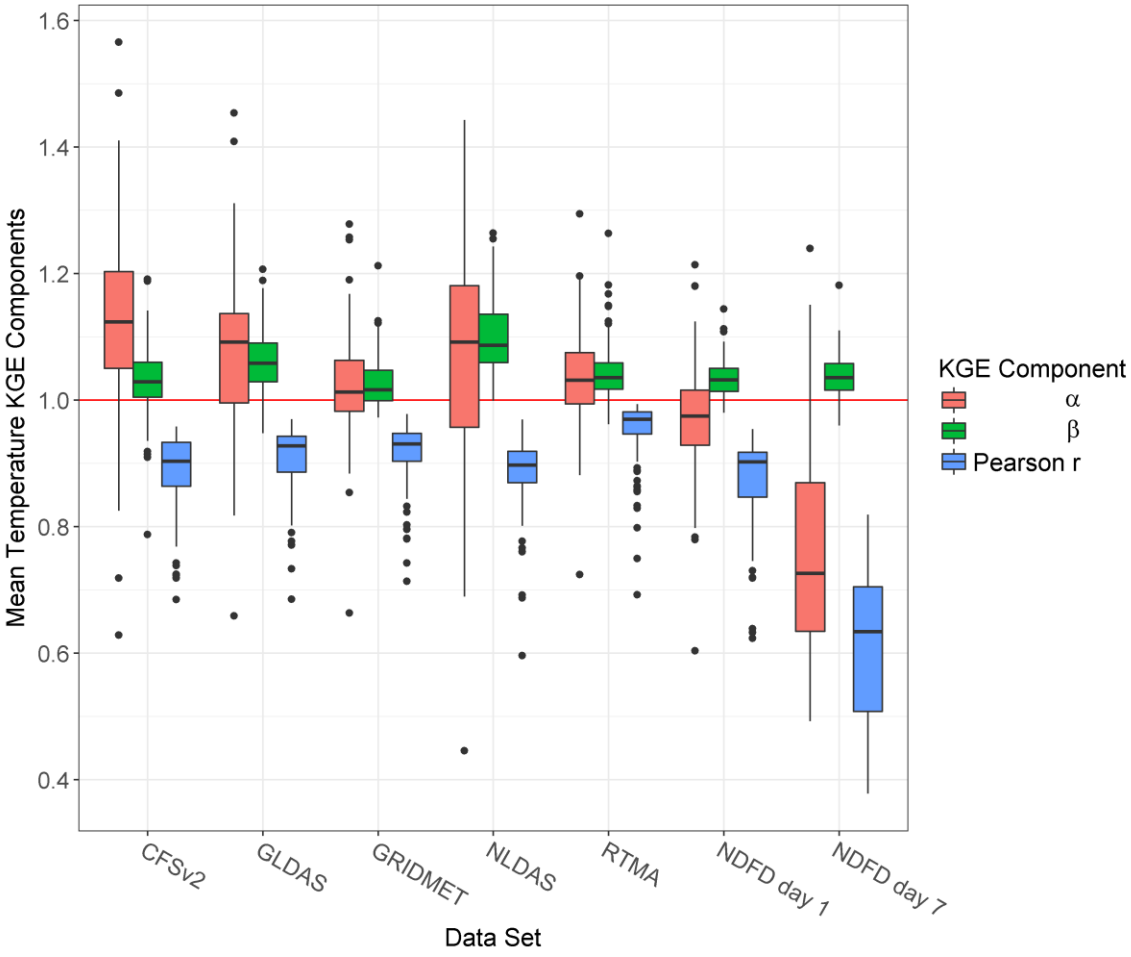

Figure 5.2 Near-surface air temperature KGE components computed using daily data at each station for the summer of 2015.

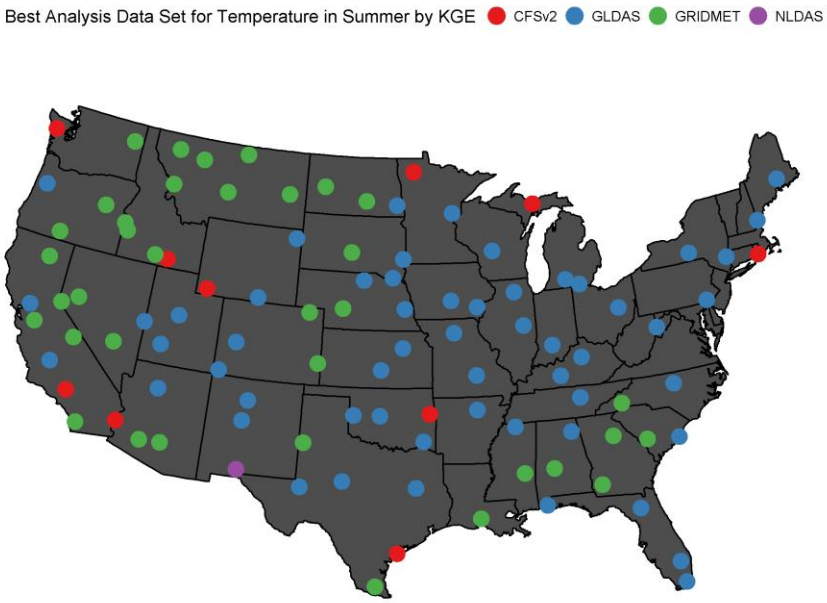

Figure 5.3 Top analysis near-surface air temperature data sets ranked by KGE computed using summertime daily data pooled across all study years.

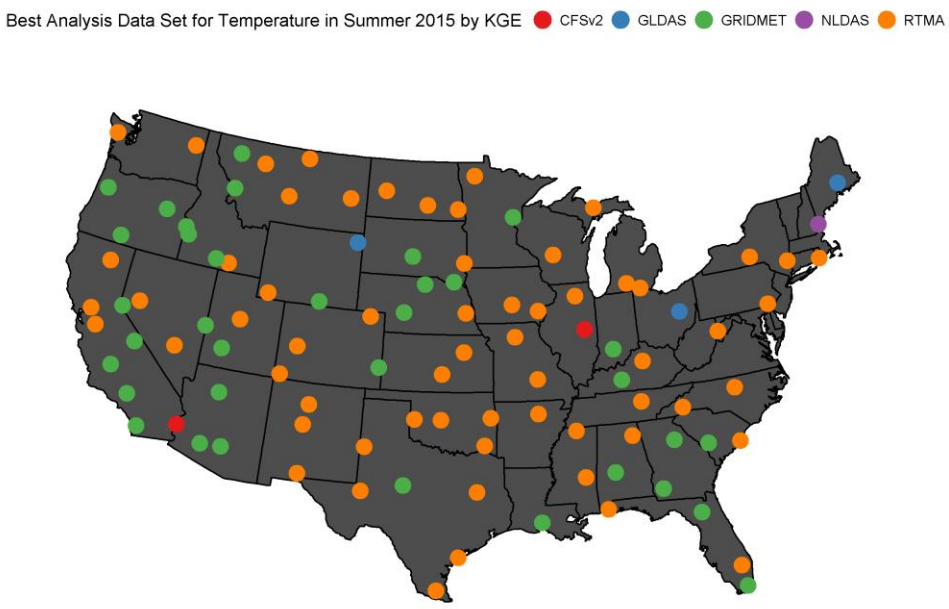

Figure 5.4 Top analysis near-surface air temperature data sets ranked by KGE computed using daily data from summer 2015.

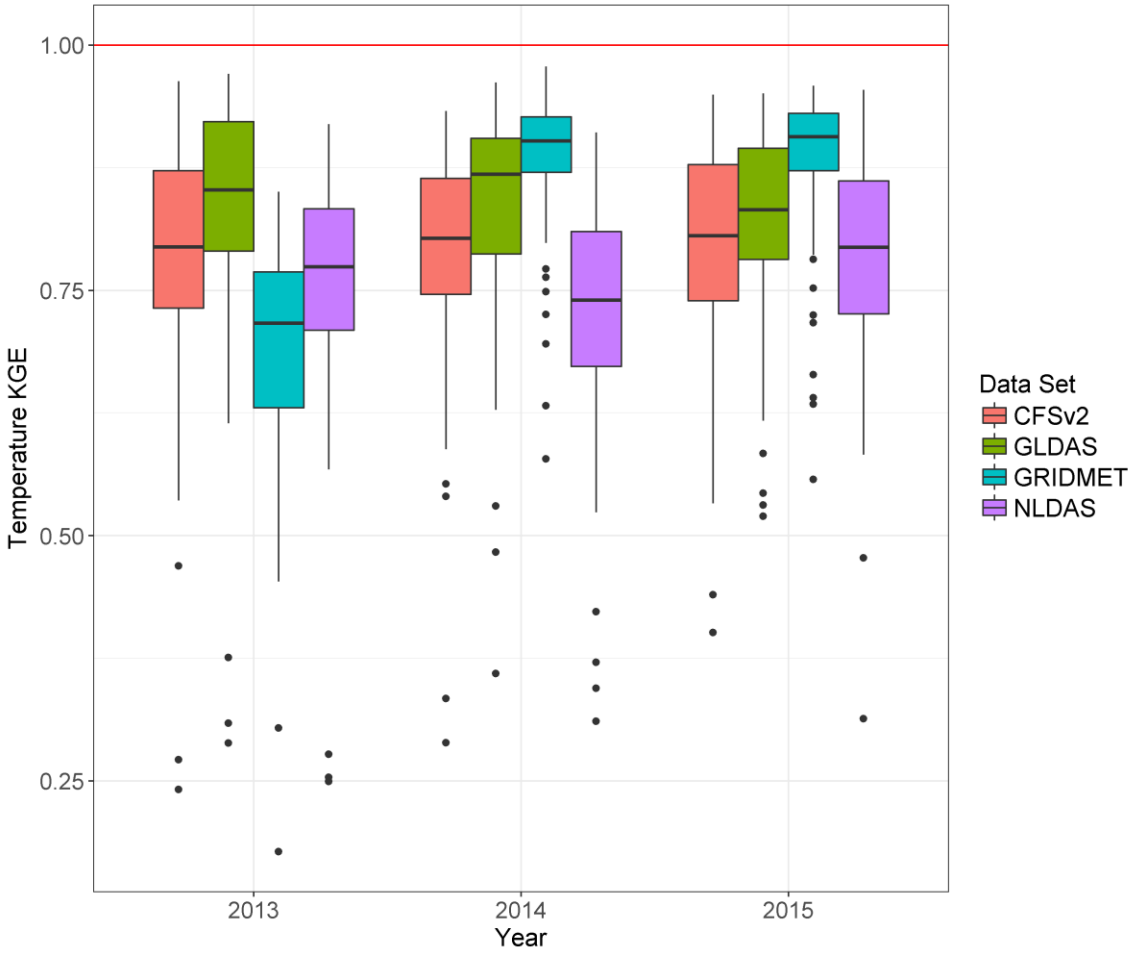

Figure 5.5 Near-surface air temperature KGE computed for the summer of each study year.

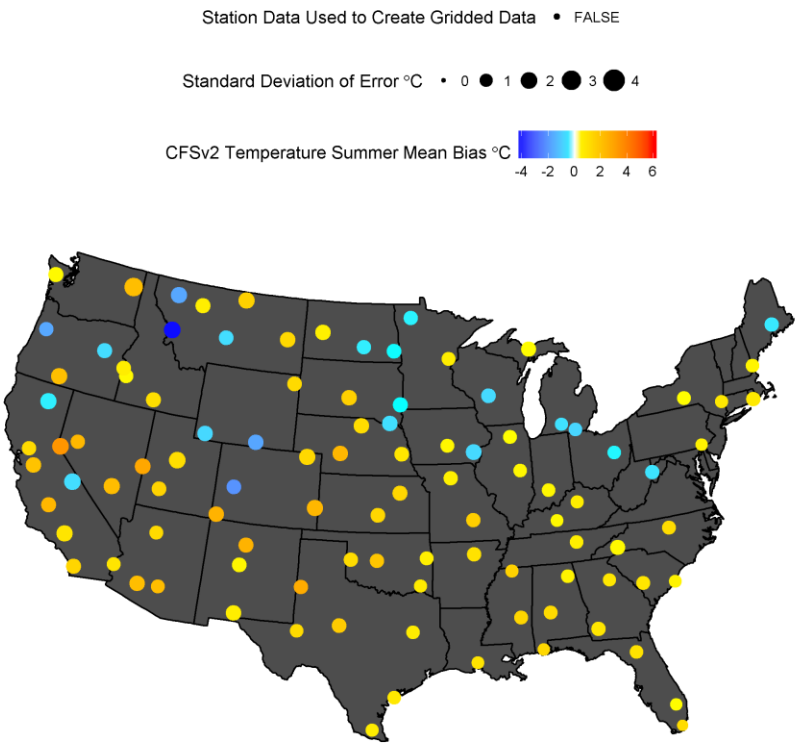

Figure 5.6 CFSv2 near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

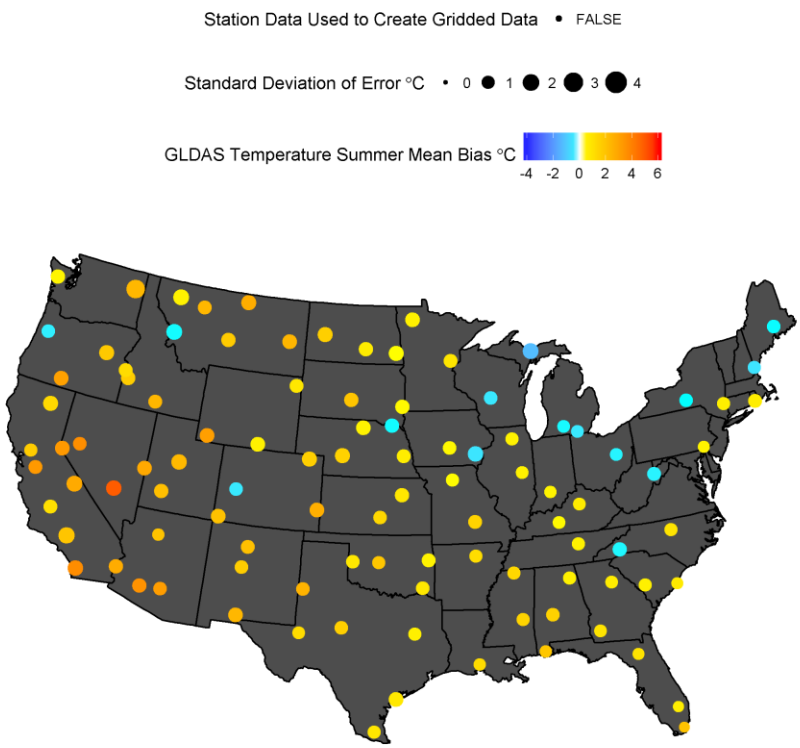

Figure 5.7 GLDAS near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

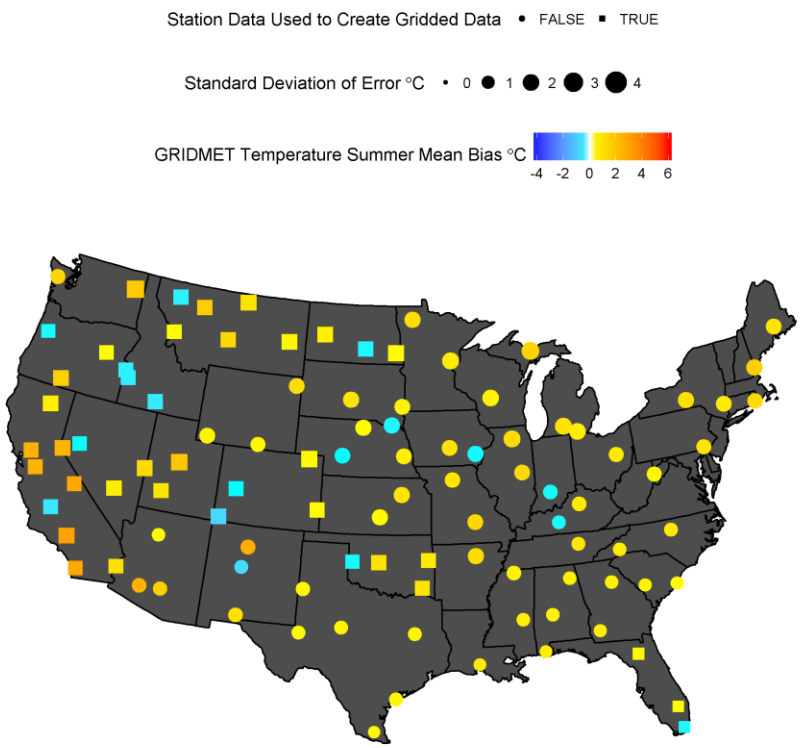

Figure 5.8 GRIDMET near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

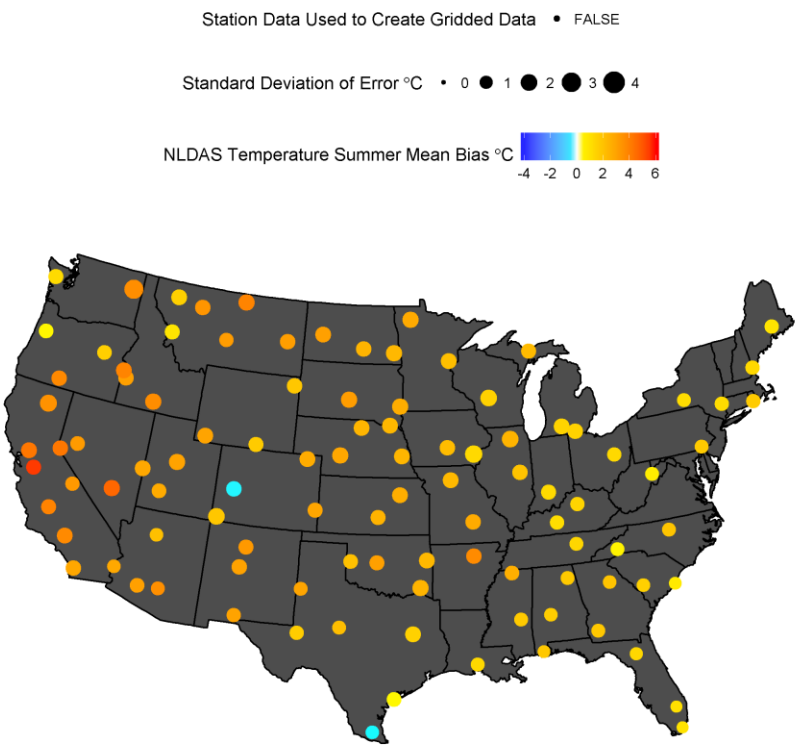

Figure 5.9 NLDAS near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

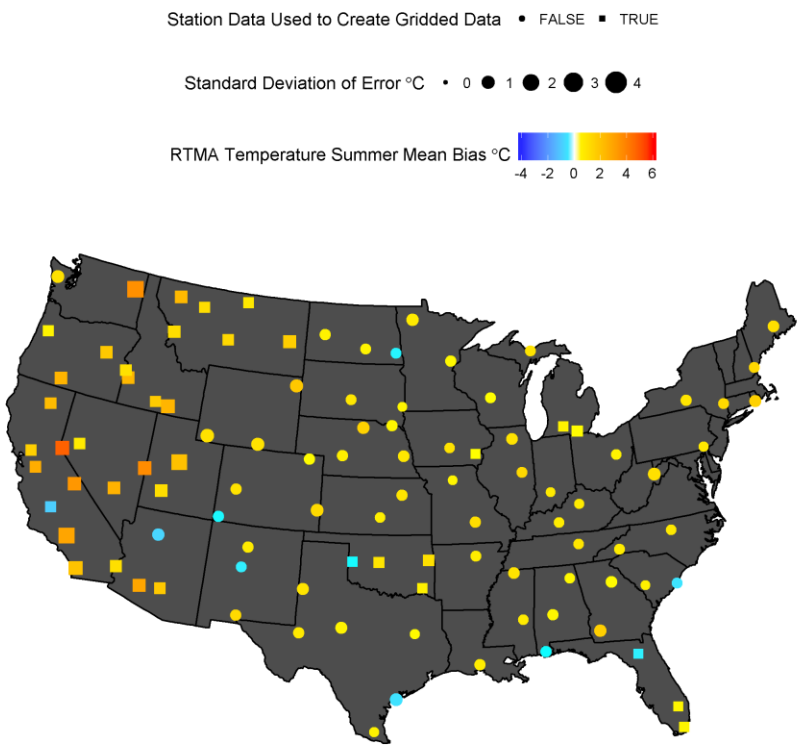

Figure 5.10 RTMA near-surface air temperature mean bias and standard deviation of error computed using summertime daily data for 2015.

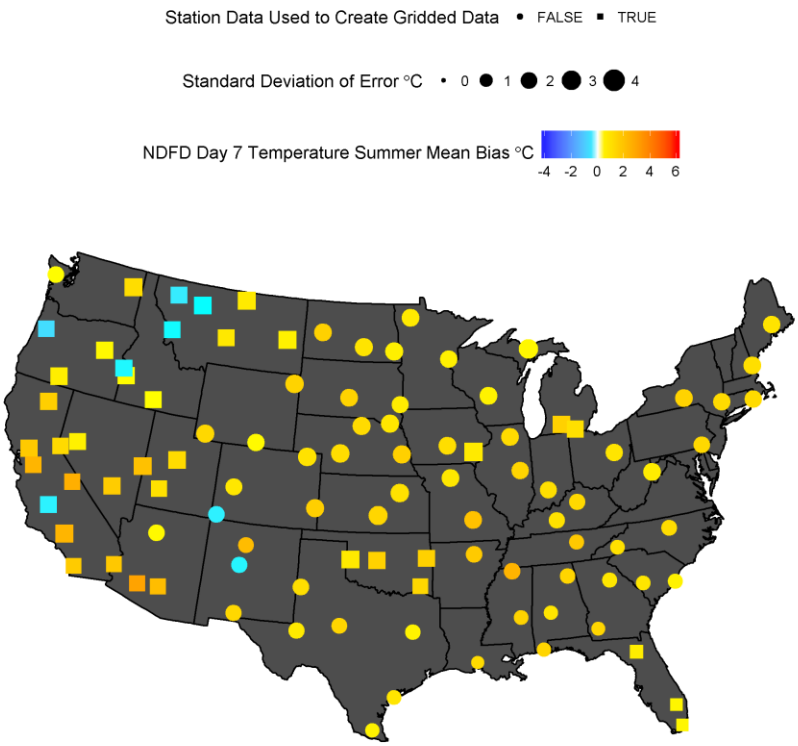

Figure 5.11 NDFD day 7 near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

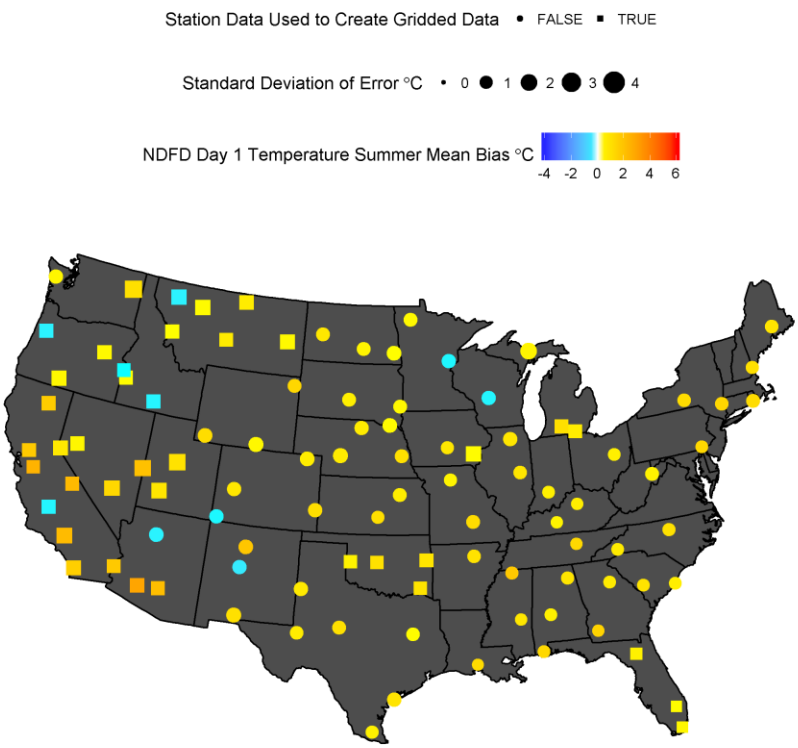

Figure 5.12 NDFD day 1 near-surface air temperature mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

5.1.2 Vapor Pressure

Vapor pressure bias was negative for more stations than it was positive for all gridded data sets, again confirming what was expected based on the literature. Maps of bias show that bias was more negative in western states. This was expected since the lower boundary layer of the atmosphere tends to gain humidity when it is contact with wetter surfaces which, in this study, existed for many of the weather station locations due to their irrigated surroundings. The gridded data sets, do not account for the irrigation and its effect on humidity and therefore underestimate it. For NLDAS, bias tended to be slightly positive in the eastern states (see Appendix D). Median biases ranged from

negative 6.2% for GRIDMET to 0.0% for RTMA. Median correlations ranged from 0.82 for GRIDMET (or 0.61 for NDFD day 7 if considering forecasts) to 0.98 for RTMA.

Vapor pressure did not show clear winners among the analysis data sets (if we exclude RTMA) (Figure 5.13). GLDAS performed best for 2013 and 2014, with its median above or near the top of the other data sets' third quartile (Figure 5.17). In 2015 the distributions overlapped more significantly. KGE rank maps show that GLDAS had the best KGE for a relative majority of stations, followed by NLDAS. It is interesting to observe that the procedure used to debias NLDAS humidity in GRIDMET does not seem to have improved GRIDMET's performance over NLDAS.

RTMA generally had the best performance among all gridded data sets for all KGE components in the summer of 2015. The rank maps show that RTMA had the best KGE for the majority of the stations. Despite RTMA's accuracy, the bias map (Figure 8.22 in Appendix D) shows that RTMA had a slight positive bias in the East and a negative bias in the West, the same spatial pattern present in the other data sets.

NDFD, likely benefiting from its relationship to RTMA, performed better than the analysis data sets at the 1-day lead time. The spatial pattern in the bias was also present in NDFD.

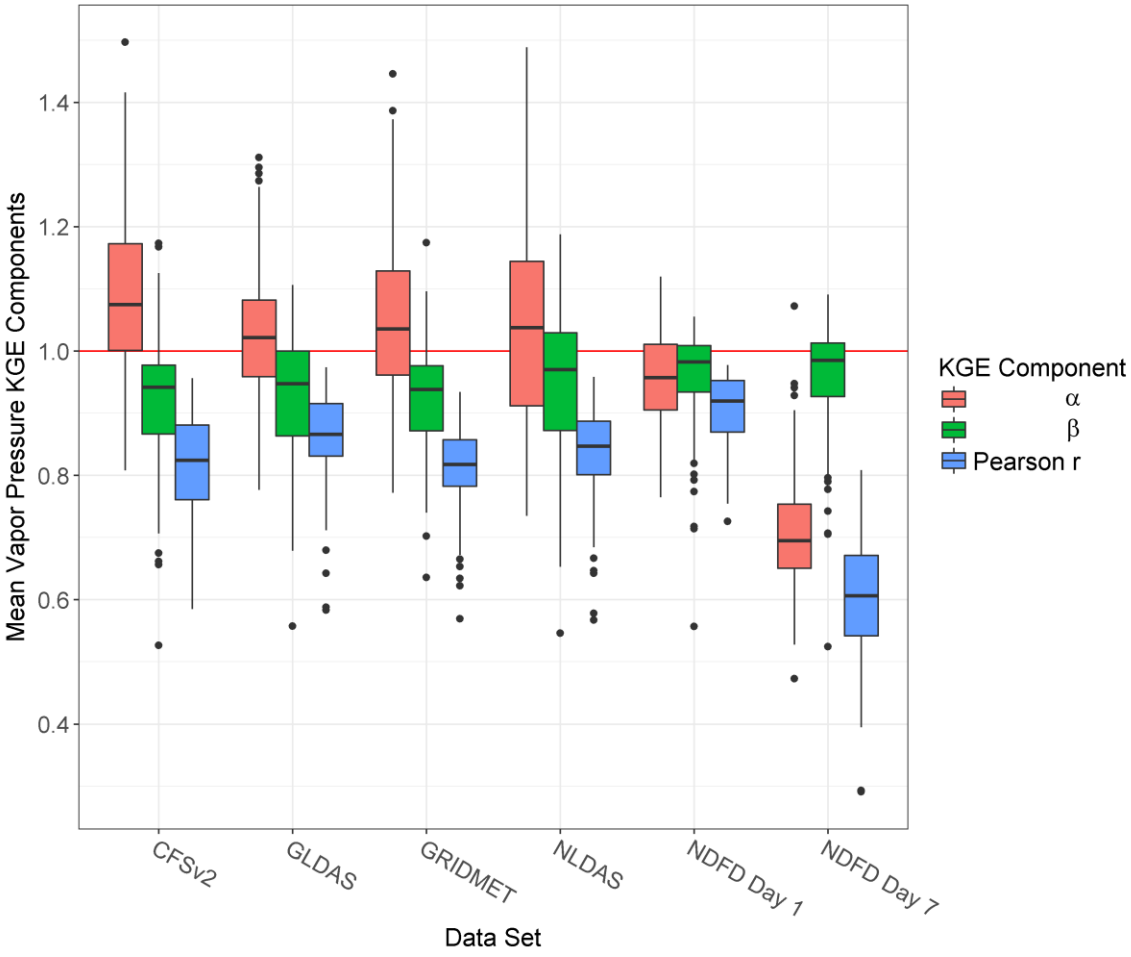

Figure 5.13 Vapor Pressure KGE components computed using summertime daily data at each station pooled across all study years.

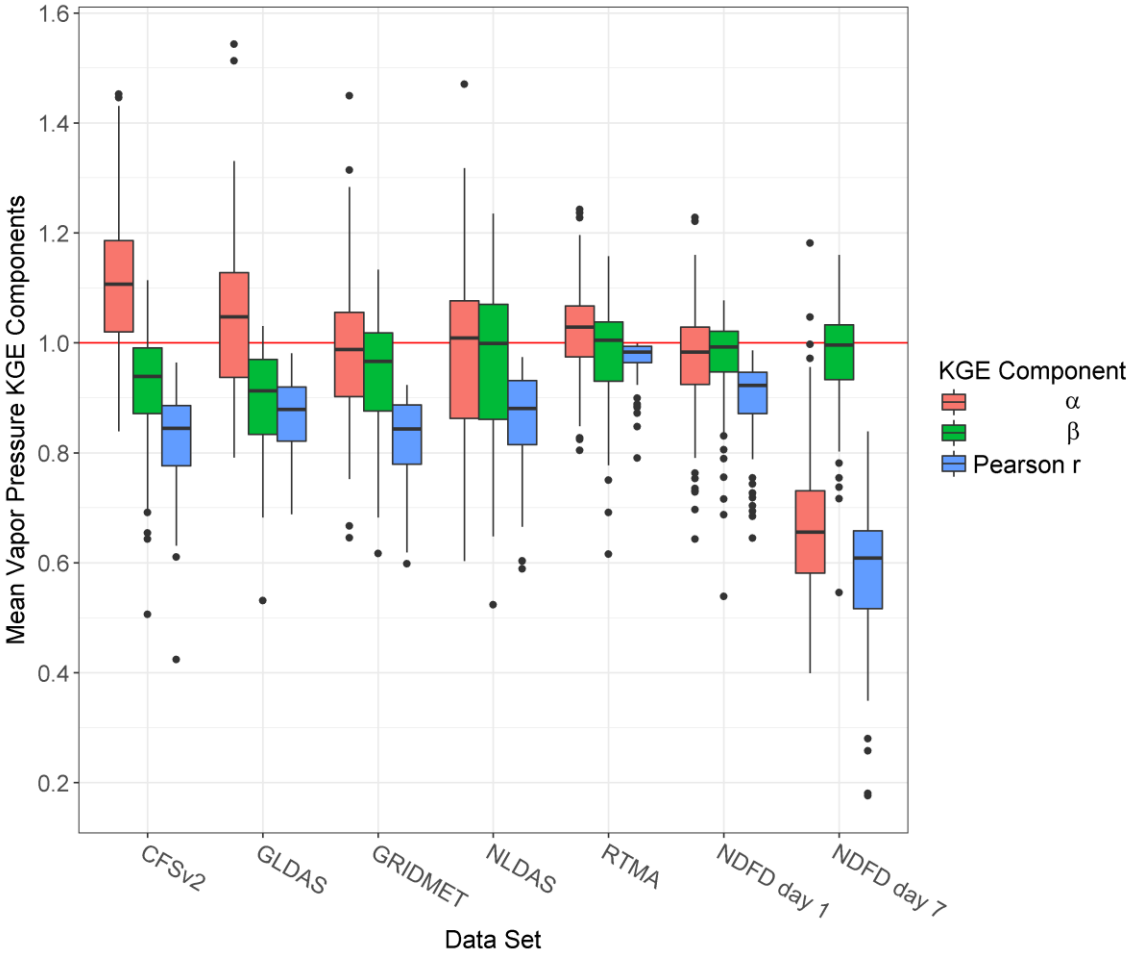

Figure 5.14 Vapor Pressure KGE components computed using daily data at each station for the summer of 2015.

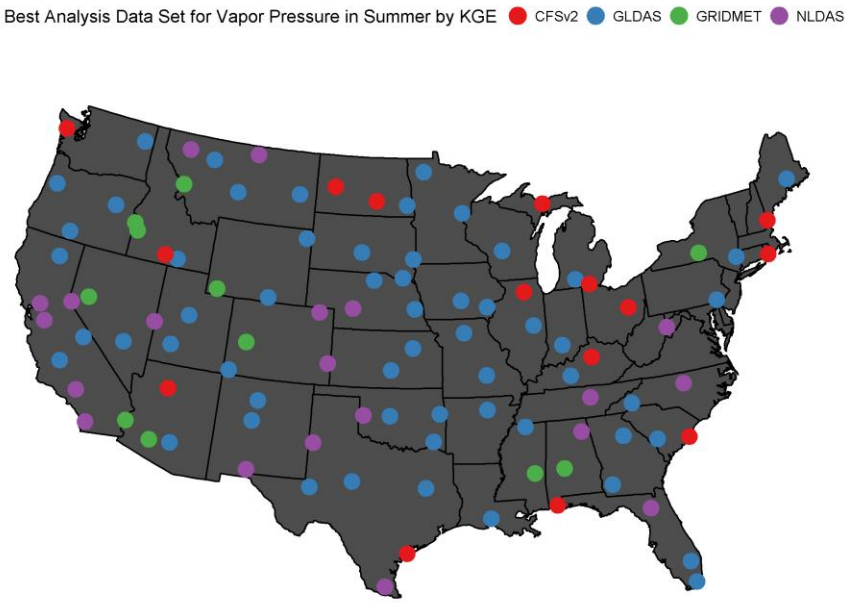

Figure 5.15 Top analysis vapor pressure data sets ranked by KGE computed using summertime daily data pooled across all study years.

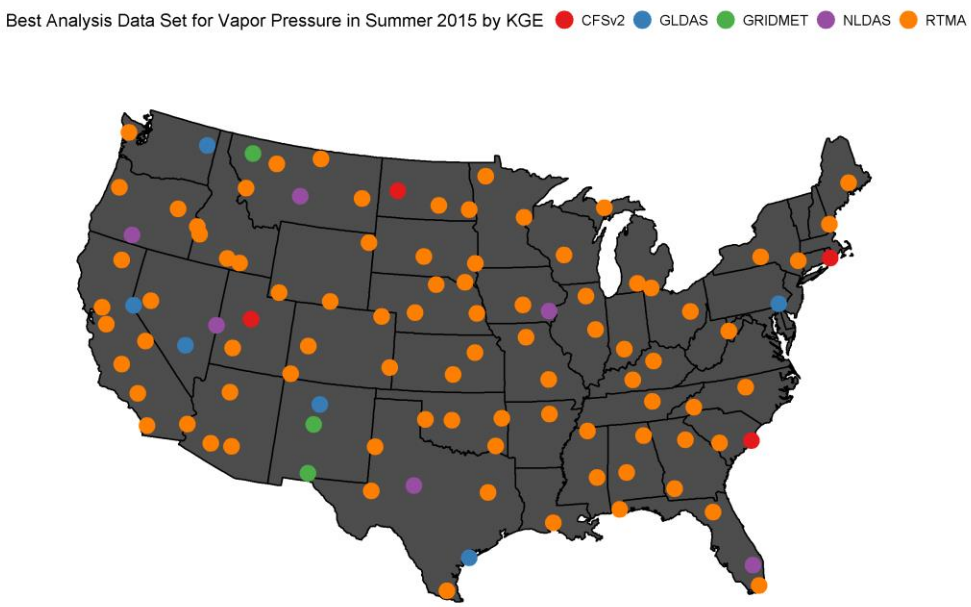

Figure 5.16 Top analysis vapor pressure data sets ranked by KGE computed using daily data from summer 2015.

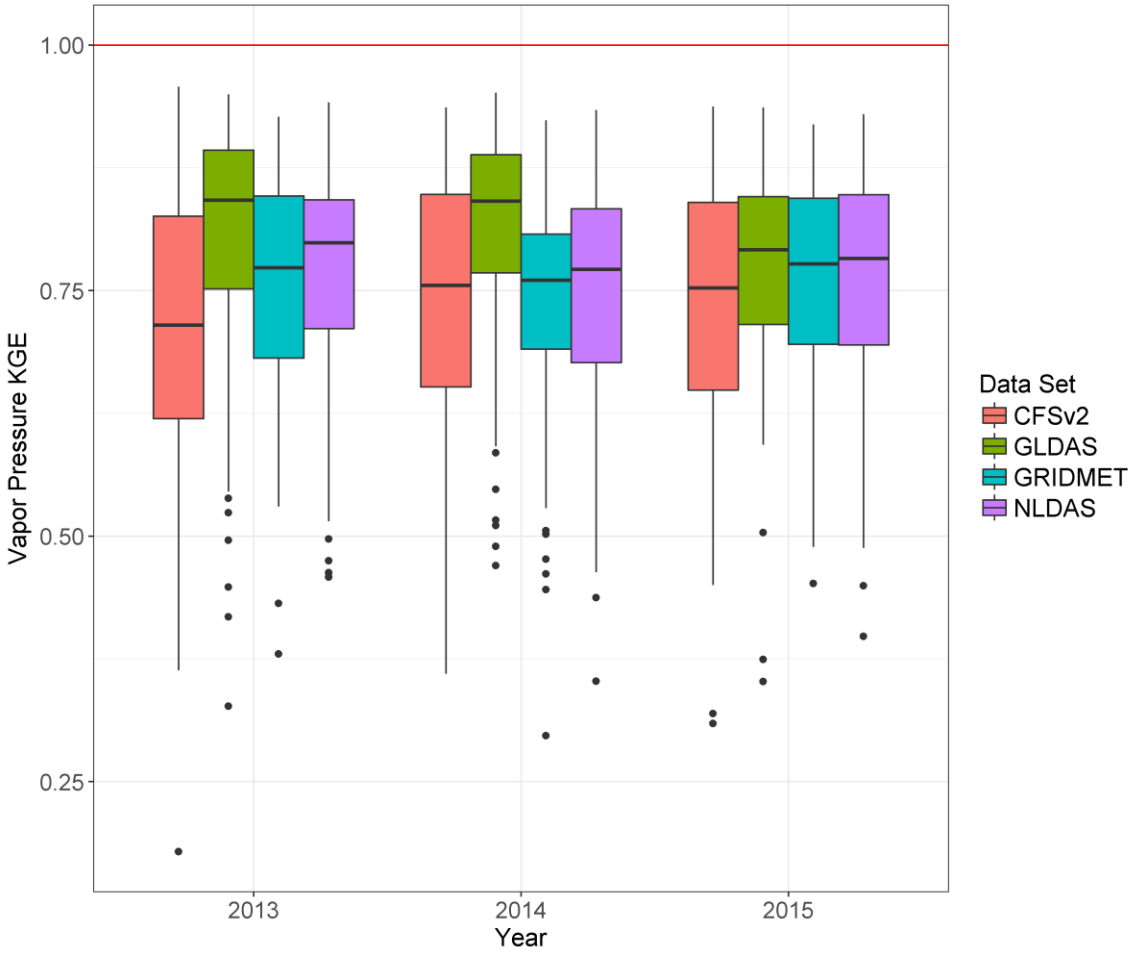

Figure 5.17 Vapor Pressure KGE computed for the summer of each study year.

5.1.3 Wind Speed

The wind speed bias was generally positive for all gridded data, and ranged from a median bias of positive 13.0% for CFSv2 to positive 41.1% for GRIDMET. CFSv2, NLDAS and GRIDMET all showed higher positive wind speed biases in the eastern states (see Appendix C). The boxplots for NLDAS and GRIDMET are extremely similar because GRIDMET only modified NLDAS by resampling it to a smaller spatial scale. The positive bias in the East is probably due to extensive forested areas that tend to shield wind speed at local weather stations. Selecting weather stations in the eastern states was

difficult because of the proximity of trees. Many stations still had trees closer than desired. Median correlations ranged from 0.68 for GRIDMET (or 0.31 for NDFD day 7 if considering forecasts) to 0.83 for RTMA. Wind speed had the lowest correlations of all the weather variables, which shows that gridded data struggle to represent wind speed on a daily timescale.

The outliers in the boxplot show some stations with very low correlations, even some negative correlations, indicating that the gridded data had little ability to represent day to day wind speeds at some locations. Among the outliers for β is a station where NLDAS and GRIDMET estimated the wind speed, translated to 2 m, to be 6 times the measured average. That outlier is a station located in northeastern Washington. That station is sited reasonably far from obstructions, so the bias may be due to a faulty anemometer. The bizarrely low wind speeds at that location (averaging around 0.5 m s^{-1}) were missed during QAQC partly because they exhibited a normal diurnal cycle.

The rankings map excluding RTMA shows a mix of best data sets, underscoring how similarly they performed. NLDAS was the best data set at more stations than expected, given the boxplot. GRIDMET appears least often on the map. Figure 5.22, showing the changes in KGE by year, shows very little variation in KGE over the study years.

The boxplot for summer 2015 (Figure 5.19) shows that RTMA had smaller distributions for β and α than the other data sets. The rank map for 2015 (Figure 5.21) shows RTMA as the best performing data set at a relative majority of stations. GLDAS

was best at the next highest proportion of stations followed by CFSv2 and NLDAS. GRIDMET did not appear as the best data set at any station location in summer 2015.

NDFD bias and variability error improved from the 1-day lead time to the 7-day lead time but are balanced out by a lower correlation. The reduction in bias is an unexpected result for which we have no explanation.

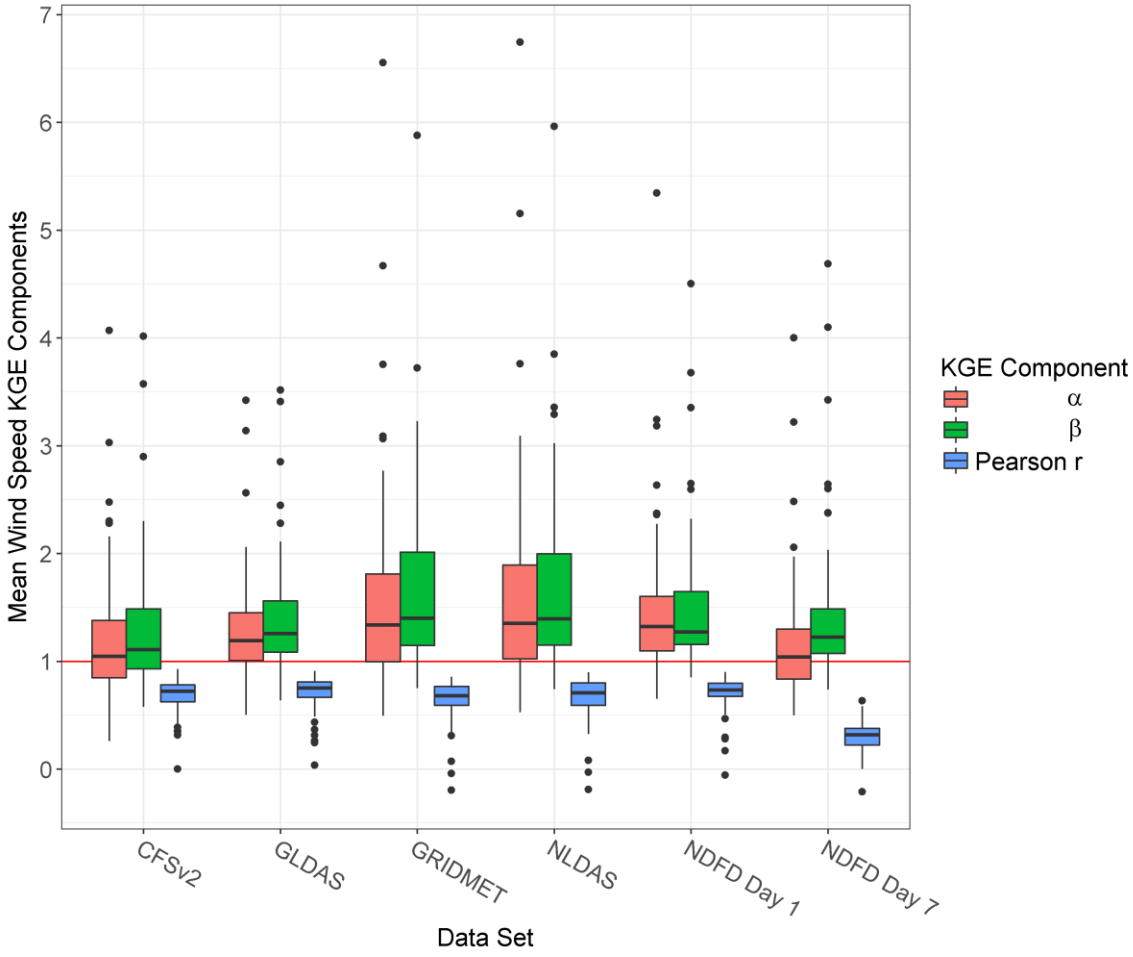

Figure 5.18 Wind speed KGE components computed using summertime daily data at each station pooled across all study years.

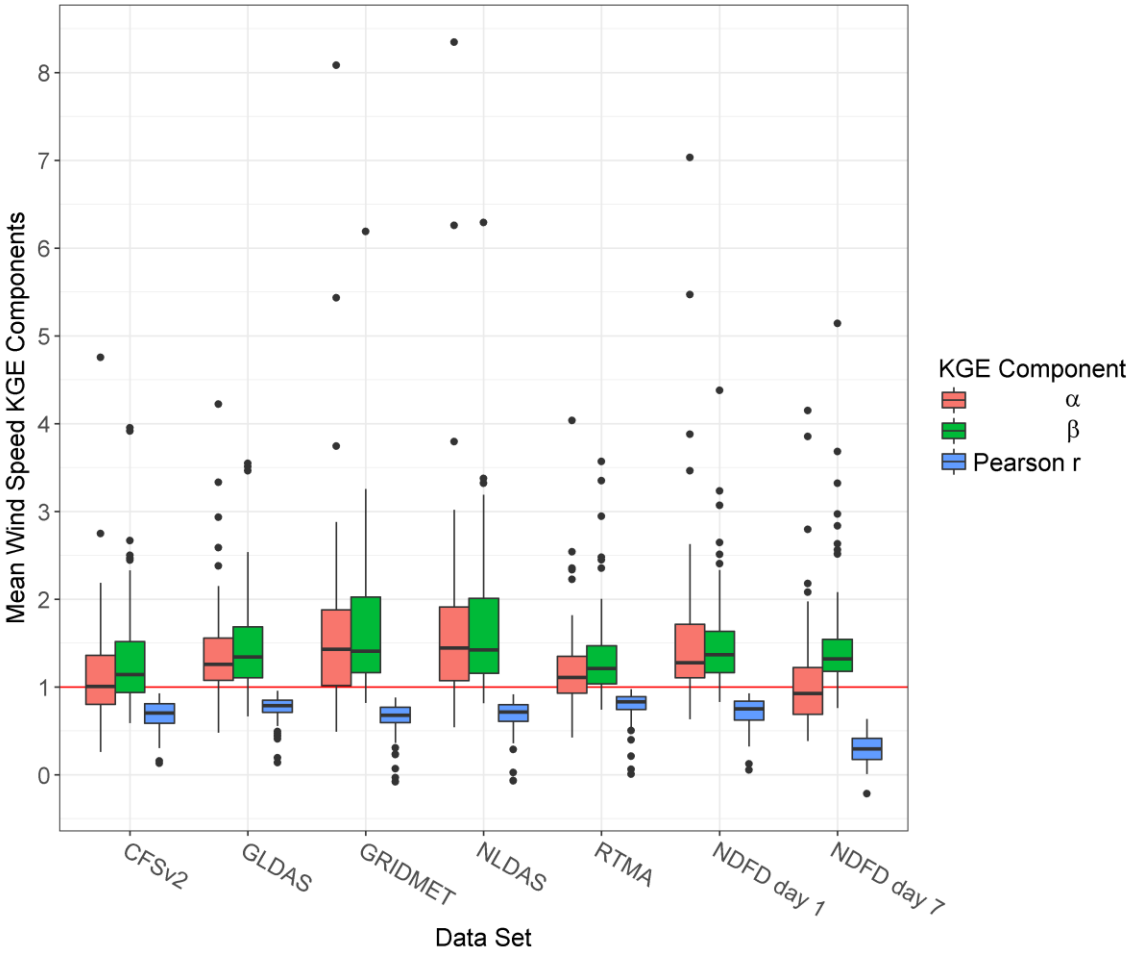

Figure 5.19 Wind speed KGE components computed using daily data at each station for the summer of 2015.

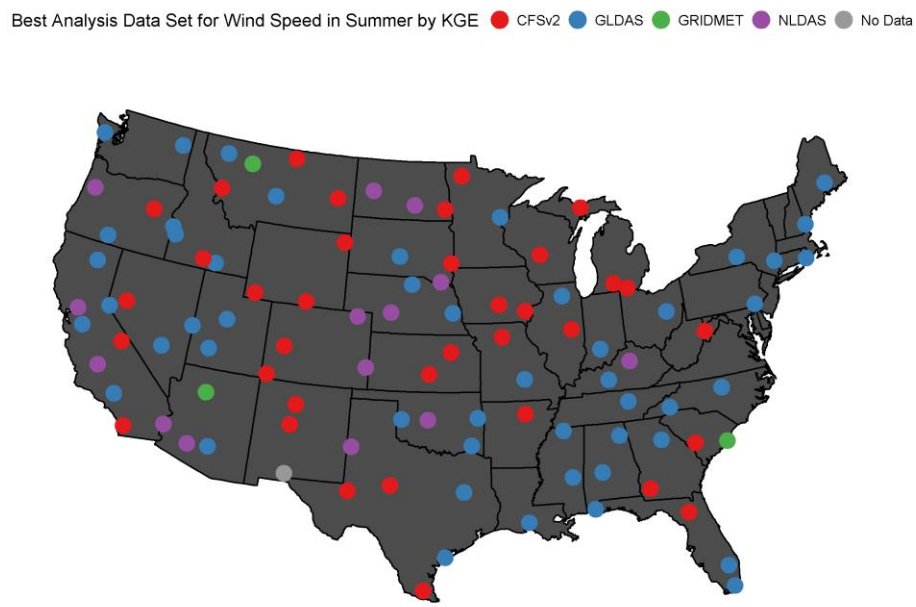

Figure 5.20 Top analysis wind speed data sets ranked by KGE computed using summertime daily data pooled across all study years.

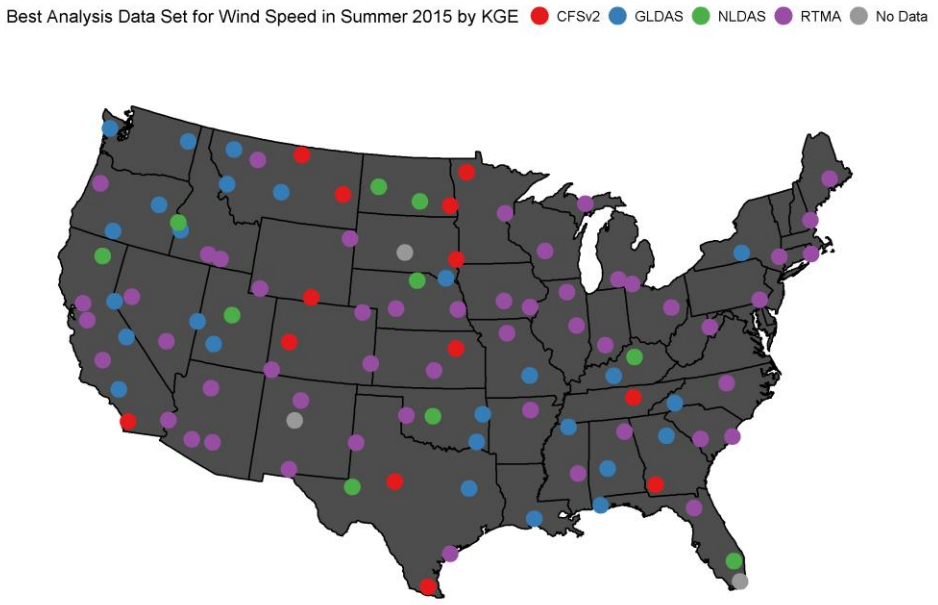

Figure 5.21 Top analysis wind speed data sets ranked by KGE computed using daily data from summer 2015.

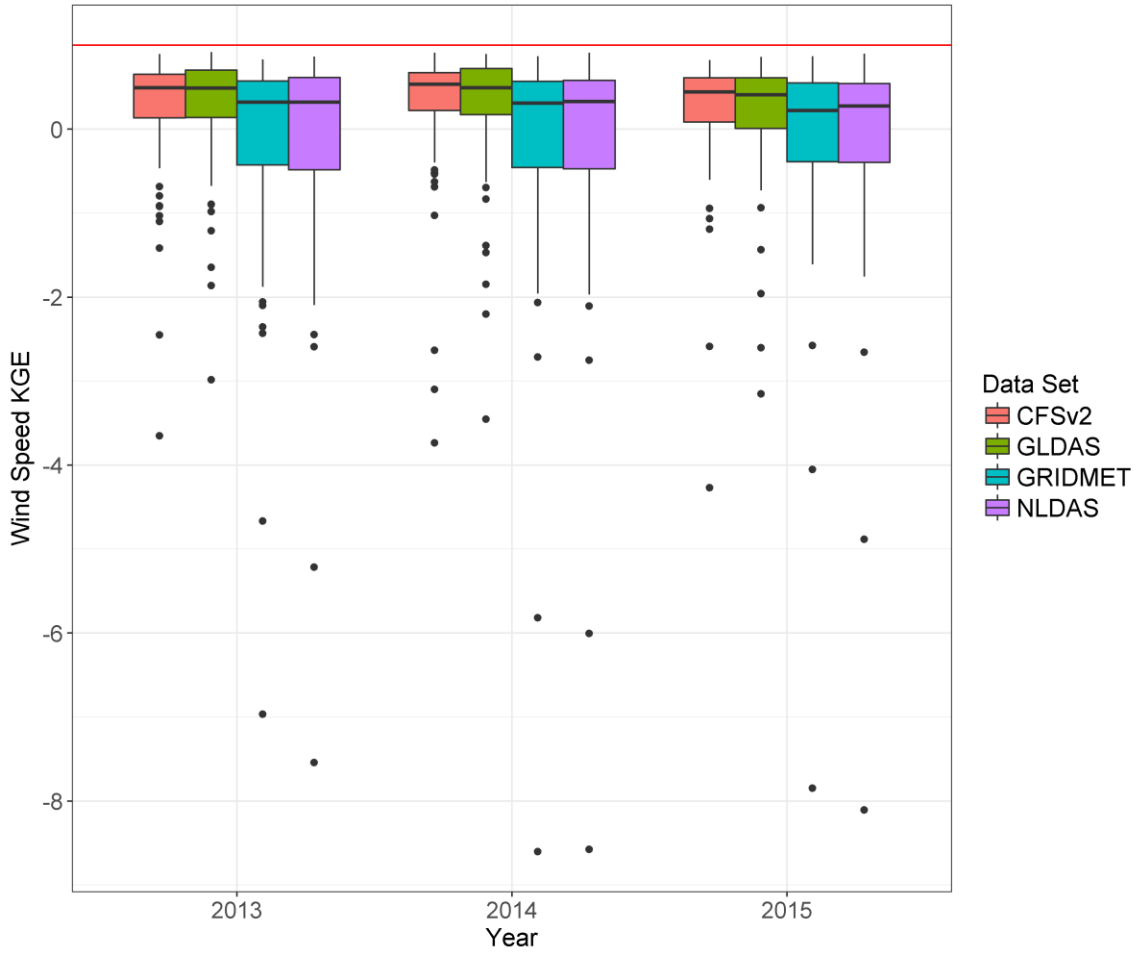

Figure 5.22 Wind speed KGE computed for the summer of each study year.

5.1.4 Solar Radiation

Solar radiation was positively biased for all analysis data sets at most stations. The median bias ranged from positive 9.9% in RTMA (or positive 0.3% including NDFD) to positive 12.1% in GLDAS. For NLDAS and GRIDMET, Utah, Nevada, western Colorado and northern Arizona had slightly negative biases (Figures 6.3 and 6.4 in Appendix B). GLDAS had smaller but still positive bias in that same area. The spatial pattern in bias matched up well with the patterns shown in Slater (2015) for NLDAS, GLDAS and CFSv2. One point of the departure in the spatial patterns is visible in GLDAS, where the Northwest had a small negative bias in Slater contrasting with the

strongly positive bias visible in our map. Another difference between the results of Slater and our results is the scale of the bias; we have generally higher positive biases than they report. This may be due to differing methods of correcting station data.

Slater corrected daily solar radiation data but we corrected the hourly data. Correcting hourly data let reductions in radiation at dawn and dusk due to topography or vegetation shading persist. Letting those reductions remain reduced our average daily solar radiation estimates relative to Slater's estimates, hence the larger biases present in our study. Slater's study may therefore be a better account of the gridded data's true error, that is, its difference from what it attempts to represent. Nevertheless, our study better represents the radiation actually impacting each site, and therefore, the error relative to that reality.

The boxplot for all study years (Figure 5.23) shows that GRIDMET and NLDAS underestimated the variability at nearly every station. Like wind speed, GRIDMET merely resamples NLDAS to a finer grid; thus the results for the two data sets are very similar. GLDAS underestimated variability at more stations than it overestimated it. CFSv2 was more evenly split between overestimation and underestimation of variability but overestimates at more stations. The results from Decker et al. (2012) suggest that both CFSv2 and GLDAS should over estimate variability more often than they under estimate it. This conflicts with our findings for GLDAS, but it could be explained by station selections that differing between our studies.

As expected, given the high variability error for GRIDMET and NLDAS, the rankings map for all study years mainly shows a blend of GLDAS and CFSv2.

Examining the boxplot for the summer of 2015 (Figure 5.24), we see RTMA had positive bias that was close to the other gridded data but a better correlation. The ranking map for 2015 shows that RTMA, and in particular RTMA computed with the Perez et al. equation, was best at the majority of stations. RTMA tends to be better relative to the other gridded data in the East and less so in the West. Refitting the Perez et al. equation to data gathered in an arid western state might produce better results in that region.

The main difference between the solar radiation derived from Perez et al. (2007) and the radiation derived from Doorenbos & Pruitt (1977) was the variability error. The Perez et al. (2007) method exhibited a substantial improvement in the variability error when estimating solar radiation for RTMA or NDFD. This is visible in both box plots (Figure 5.23 and Figure 5.24). The linearity of the Doorenbos & Pruitt function caused it to underestimate in the western states where cloud cover is less (see Appendix B). In Doorenbos & Pruitt, they explained that the relationship between sky cover and radiation should ideally be determined with local data because "…scatter in conversion factors from location to location has been noted." Figure 5.27, a ranking map of the 1-day lead time NDFD solar radiation predicted with Doorenbos & Pruitt and Perez et al. methods, shows that the improvement in the variability error was enough to give the Perez et al. method a significant edge at most stations.

The KGE stability plot (Figure 5.28) shows that GLDAS generally performed best for all three study years (excluding RTMA). Solar radiation correlation in NLDAS was lower than shown in Lewis et al. 2011 because they did their analysis on an hourly timestep. Hourly data produce a better correlation because of the regularity of the solar

diurnal cycle; averaging solar radiation to daily values removes the diurnal cycle and makes correlations more dependent on representing cloudiness. Median correlations for solar radiation ranged from 0.76 for GRIDMET (or 0.37 for NDFD day 7 if including forecasts) to 0.91 for RTMA.

NDFD generally had less bias than the other gridded data sets including RTMA. If the use of the sky cover to solar radiation conversion equations were the cause of the small bias one would expect RTMA to mirror the bias in NDFD. This unexpected result deserves more consideration and analysis.

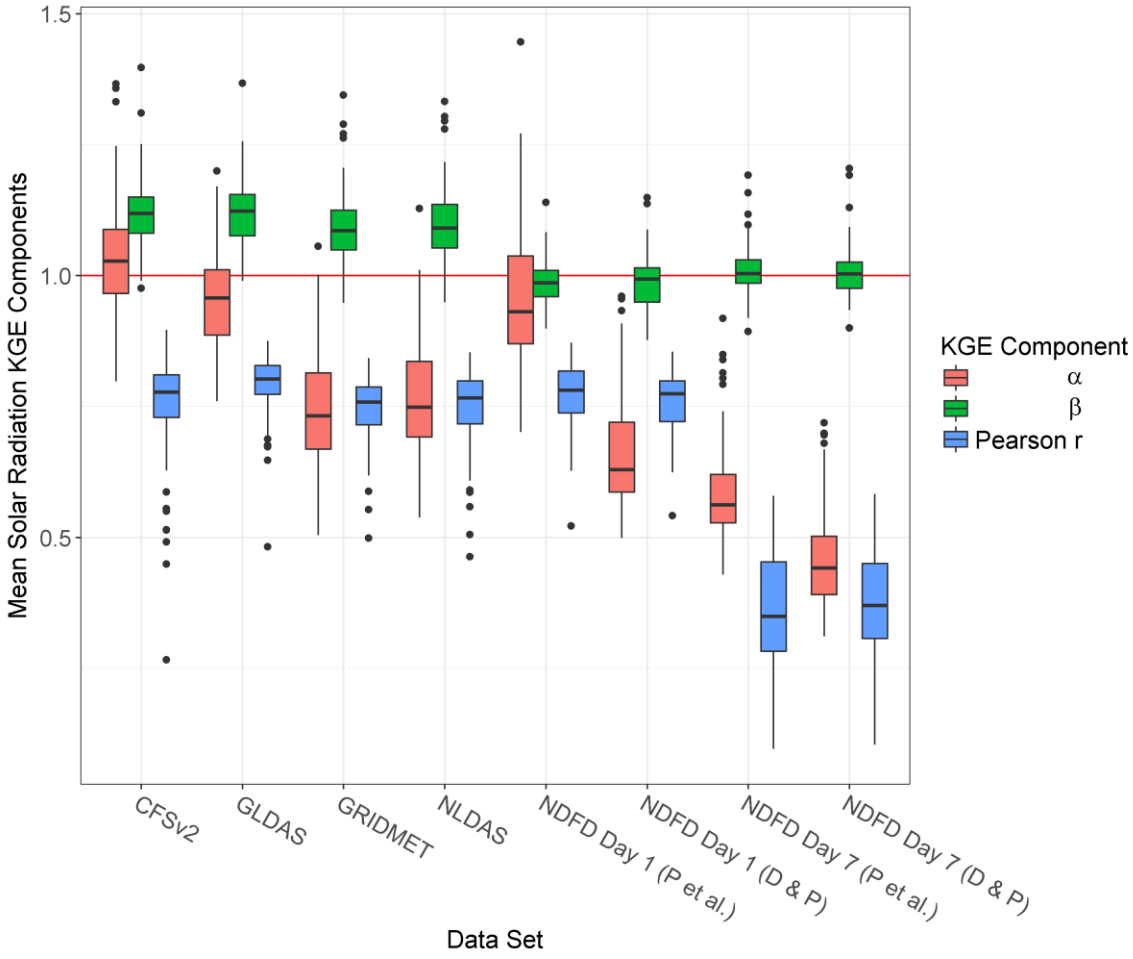

Figure 5.23 Solar radiation KGE components computed using summertime daily data at each station pooled across all study years.

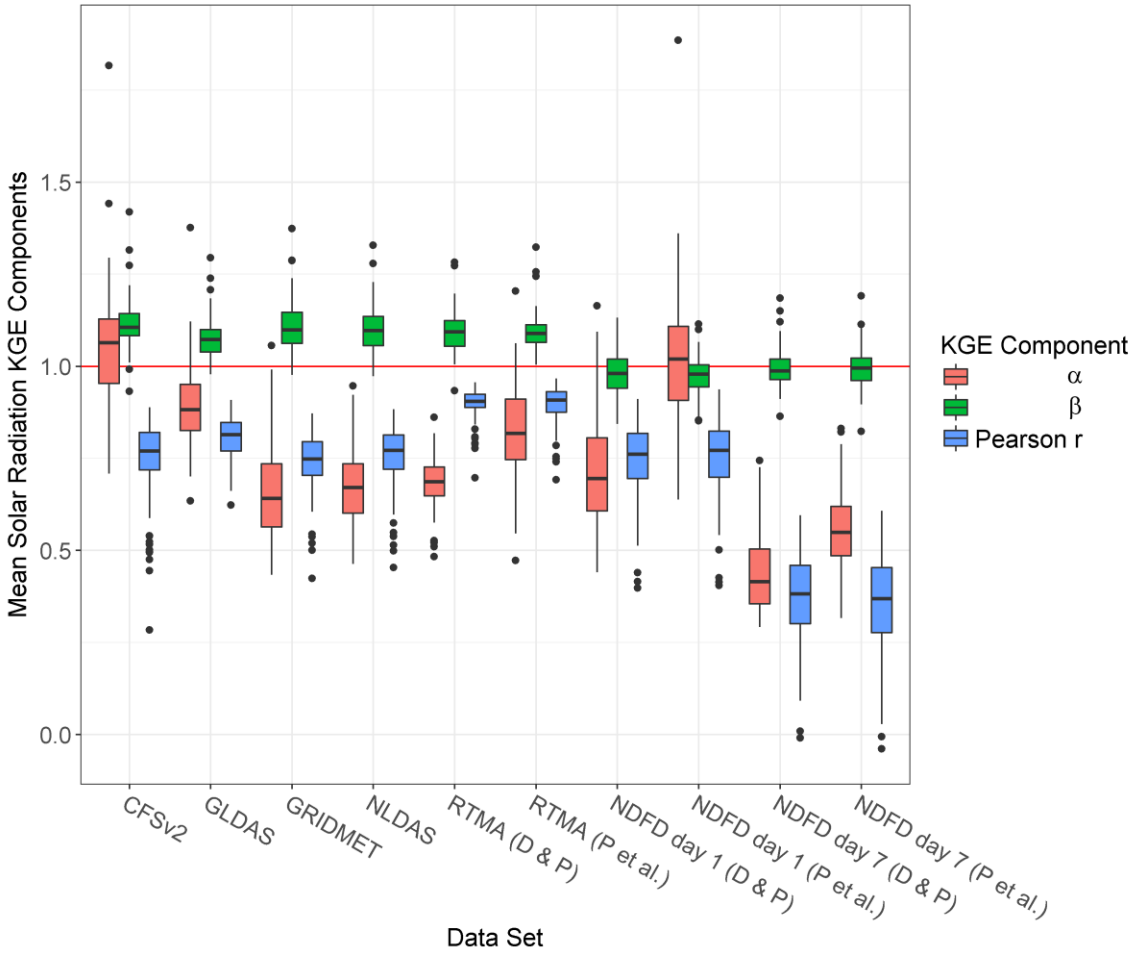

Figure 5.24 Solar radiation KGE components computed using daily data at each station for the summer of 2015.

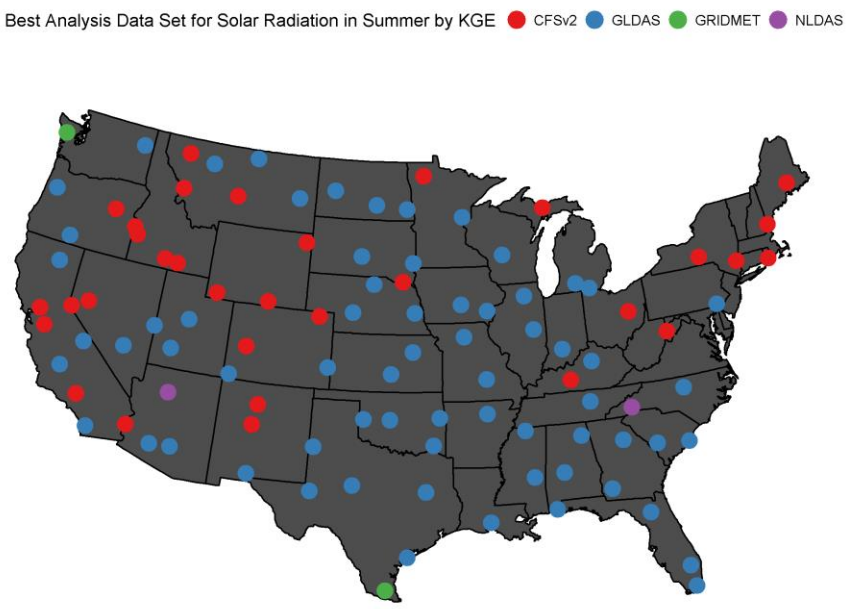

Figure 5.25 Top analysis solar radiation data sets ranked by KGE computed using summertime daily data pooled across all study years.

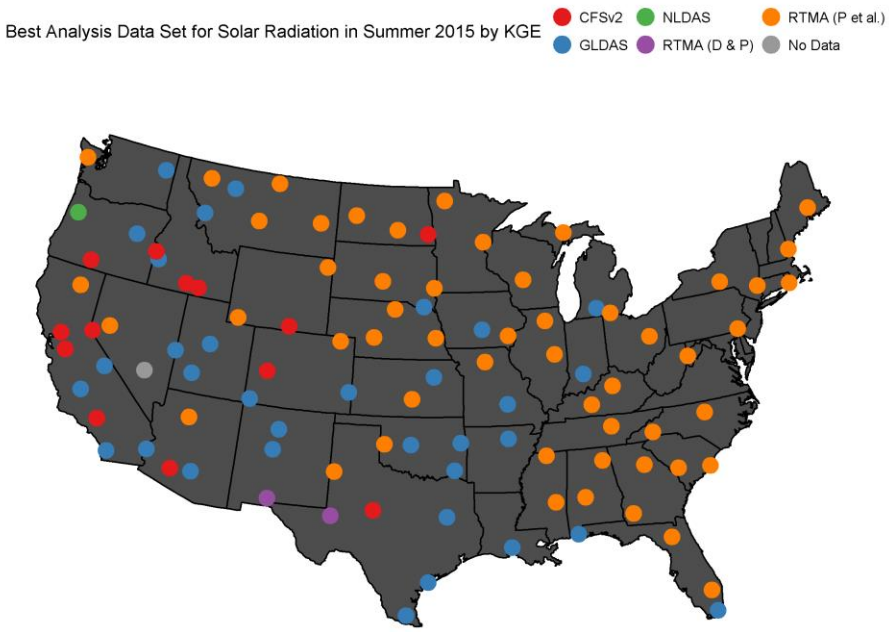

Figure 5.26 Top analysis solar radiation data sets ranked by KGE computed using daily data from summer 2015.

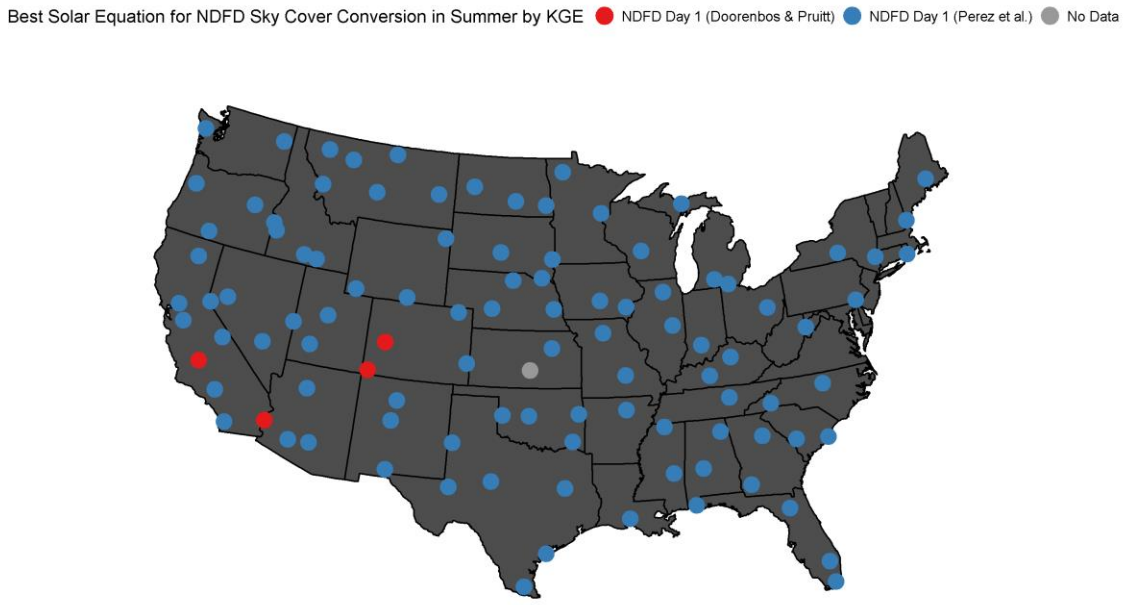

Figure 5.27 Top sky cover to solar radiation conversion equation for NDFD ranked by KGE computed using summertime daily data pooled across all study years.

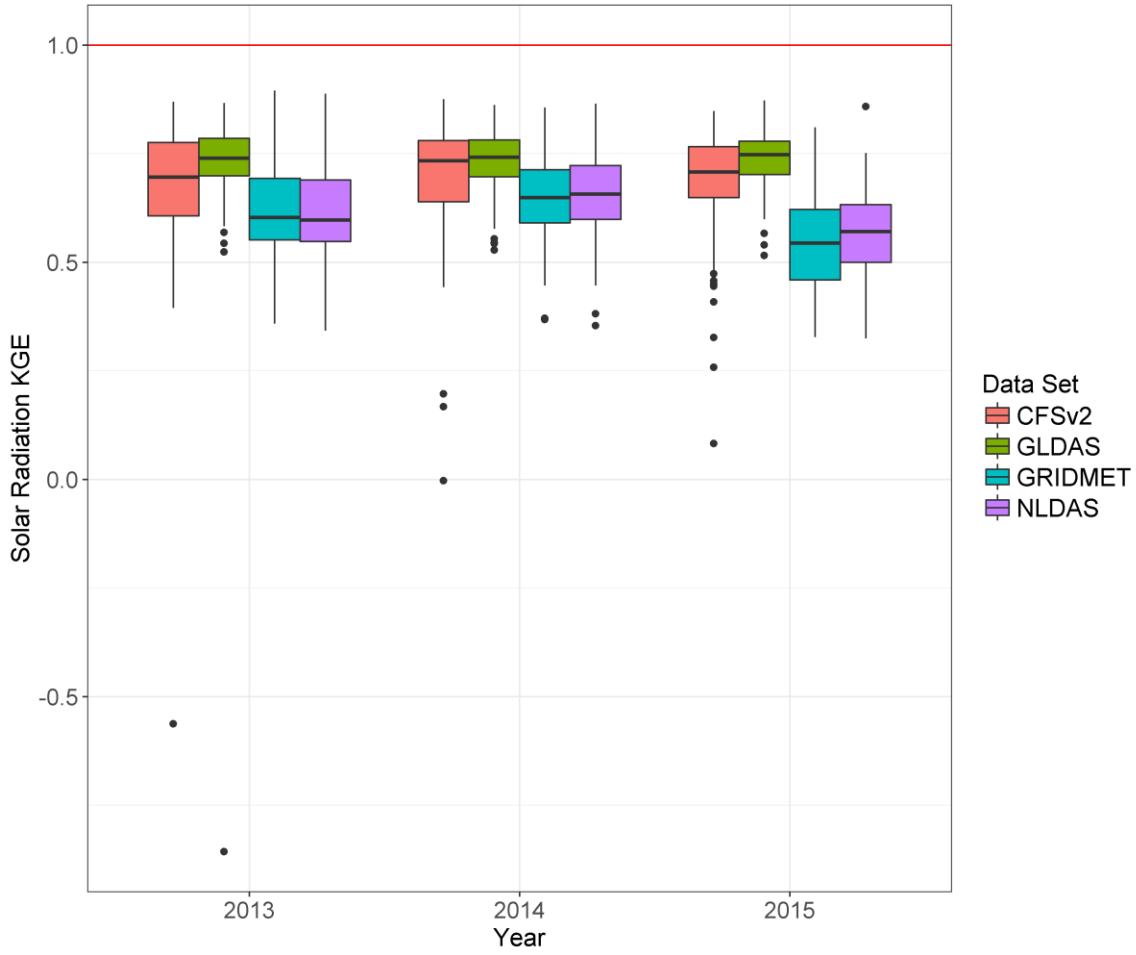

Figure 5.28 Solar radiation KGE computed for the summer of each study year.

5.2 Reference Evapotranspiration

Computed ET_r had a positive bias for all gridded data at most weather stations. Spatial patterns in the bias are difficult to discern because of the complex interplay of the error in each of the input variables. Bias does seem more volatile in the western states, with the magnitude of the bias varying widely. The standard deviation of the error in the west also seems to be more volatile, with some stations having very high standard deviations and others very small (see Figures 5.35 through 5.44). The eastern states show more uniformity in bias and the standard deviation of error. GRIDMET and NLDAS both had high positive bias in the Plains region, parts of California and the states

bordering Mexico. They have low bias in the East and in the Great Basin area. NDFD had noticeably higher positive biases in the Southwest. Median bias ranged from positive 12.5% for RTMA computed with Doorenbos & Pruitt (or 12.0% for NDFD day 1 if considering forecasts) to positive 31.2% for NLDAS. Median correlation ranges from 0.70 for NLDAS (or 0.37 for NDFD day 7 if considering forecasts) to 0.92 for RTMA computed with the Perez et al. method.

The relatively large overestimation of ET_r is a concern for use of these estimates for irrigation and water resources management. Essentially, all four weather components had biases that tend to cause ET_r to be overestimated. These include positive biases in solar radiation, air temperature and wind speed and a negative bias in vapor pressure all of which increase the computed ET_r.

According to Figure 5.29 ET_r variability was overestimated at a majority of stations for all gridded data except NDFD with a 7-day lead time and NDFD where solar radiation comes from Doorenbos & Pruitt with a 1-day lead time. Figure 5.30 shows that NLDAS and RTMA computed with Doorenbos & Pruitt equally overestimated and underestimated variability and that RTMA computed with Perez et al. overestimated the variability. This is an interesting result because it shows that using solar radiation computed using Perez et al. results in higher variability error in ET_r. The underestimation of variability by the Doorenbos & Pruitt equation might balance out an overestimation of variability in another variable (e.g. wind speed). This highlights the complex interplay between the input error when computing ET_r. A more accurate estimation of a weather variable does not necessary result in a better estimate of ET_r.

The boxplot of all study years (Figure 5.29) shows that GLDAS generally had the highest correlation and lowest variability error among the analysis data sets, but all the distributions substantially overlap. The boxplot of summer 2015 (Figure 5.30) shows that RTMA had the best performance in each KGE component for most stations.

The rank map for all study years (Figure 5.31) shows that GLDAS performed better in the eastern states. GRIDMET and CFSv2 had the best KGE at many stations in the West. GLDAS's success in the East is because of its excellent relative performance in that region estimating solar radiation and air temperature. GRIDMET's more accurate temperature allowed it to be the best data set at a surprisingly large number of stations.

Figure 5.34 shows that GLDAS ET_r performed better for 2013 and 2014 than it did for 2015. That temporal pattern aligns with the pattern in Figure 5.5 and Figure 5.17, where temperature and vapor pressure KGE for GLDAS was worse for 2015. GRIDMET's ET_r performance for 2013 was much worse than for 2014 and 2015 which aligns with its poor temperature performance for 2013.

The final rank plot (Figure 5.33) shows that NDFD had a better KGE at some locations than all the analysis data sets. This is surprising given that NDFD is a forecast, and serves as a verification of the progress made in weather forecasting.

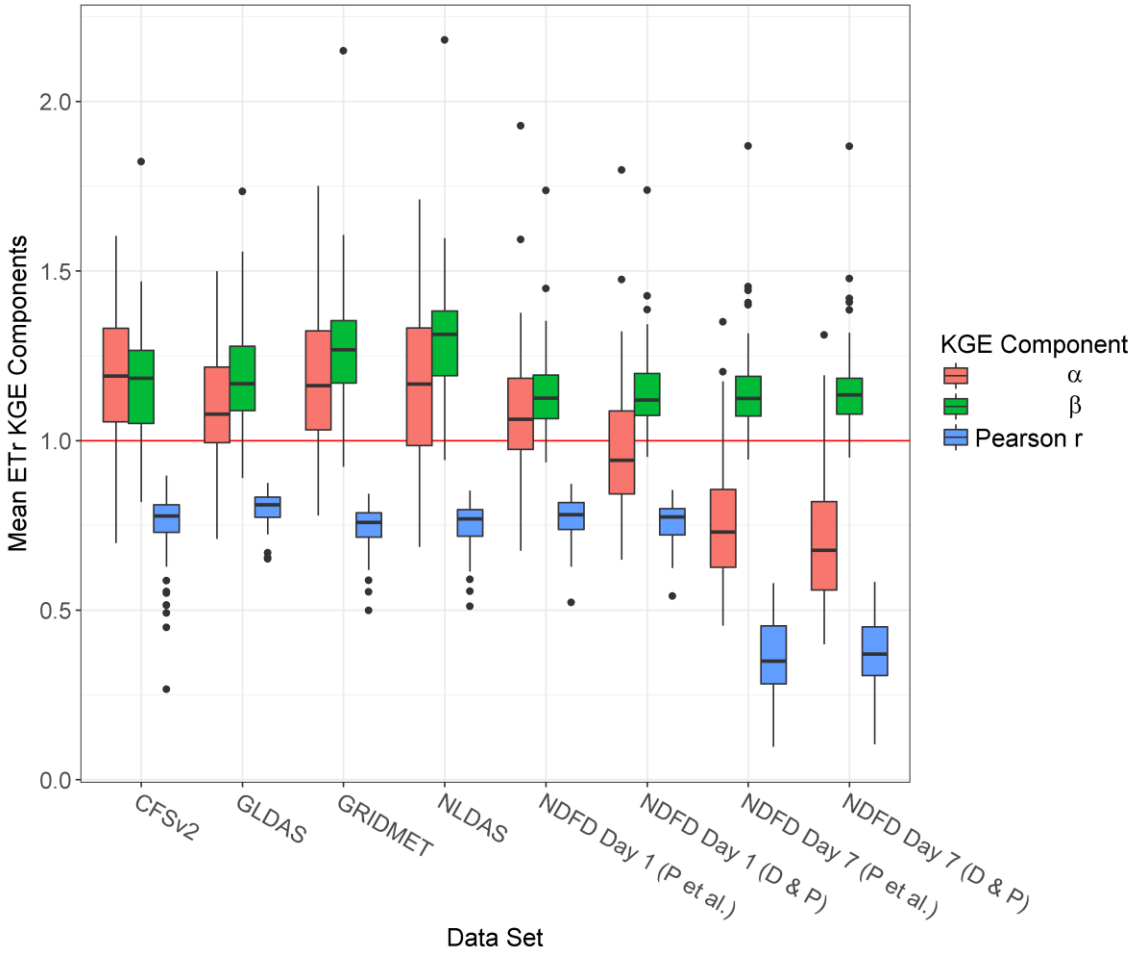

Figure 5.29 ET_r KGE components computed using summertime daily data at each station pooled across all study years.

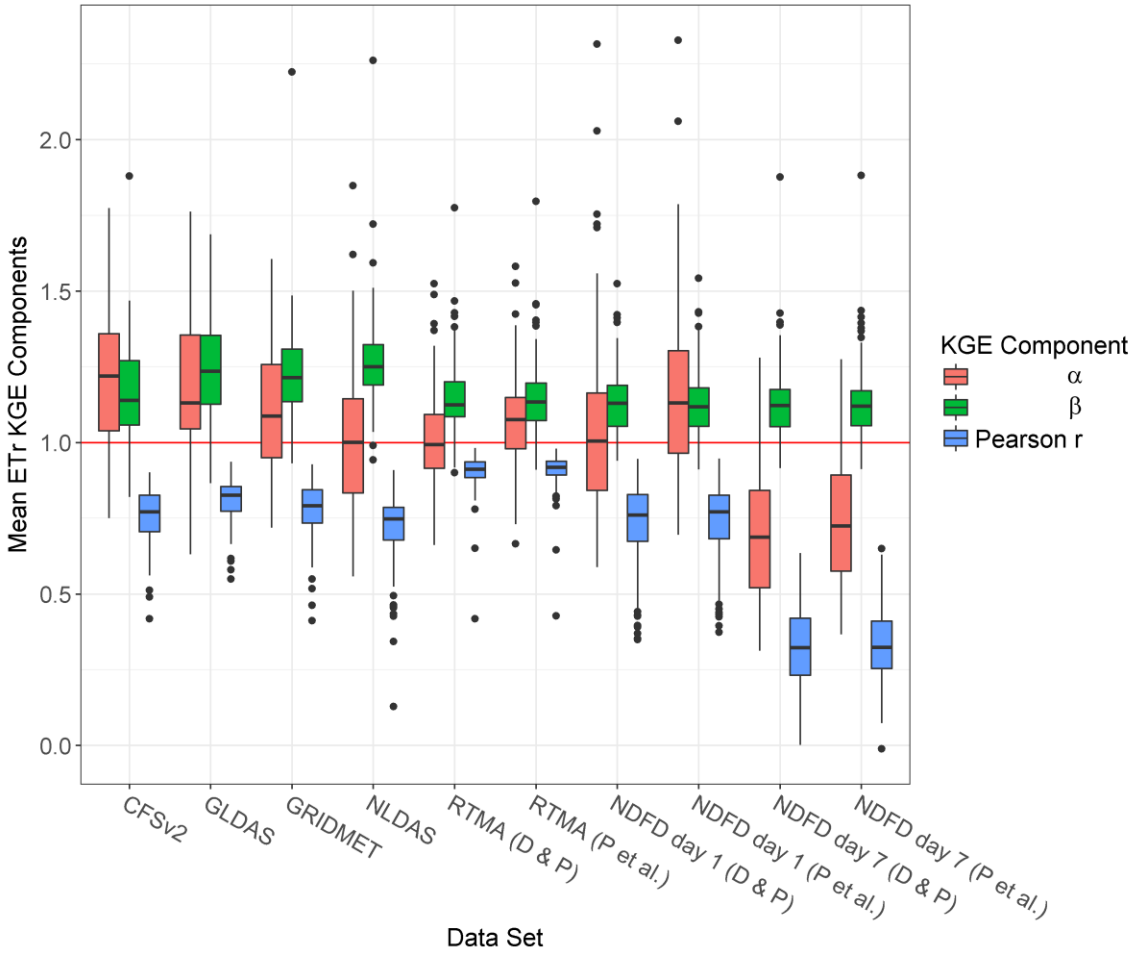

Figure 5.30 ET_r KGE components computed using daily data at each station for the summer of 2015.

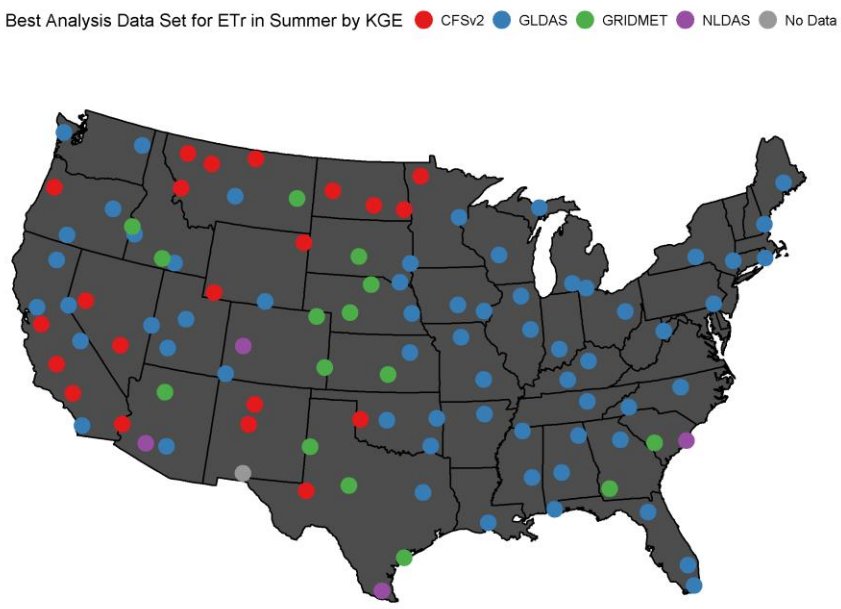

Figure 5.31 Top analysis ET_r data sets ranked by KGE computed using summertime daily data pooled across all study years.

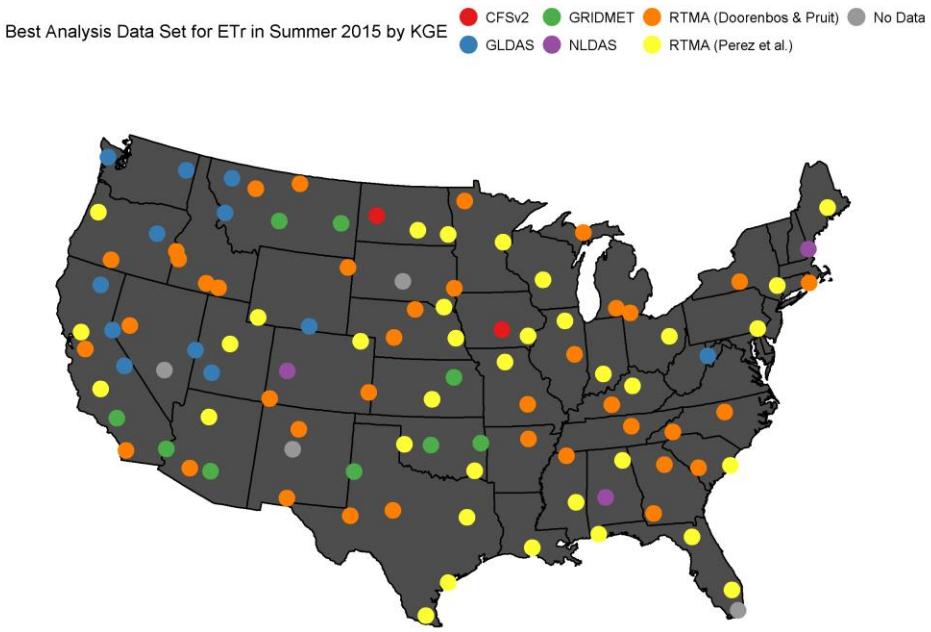

Figure 5.32 Top analysis ET_r data sets ranked by KGE computed using daily data from summer 2015.

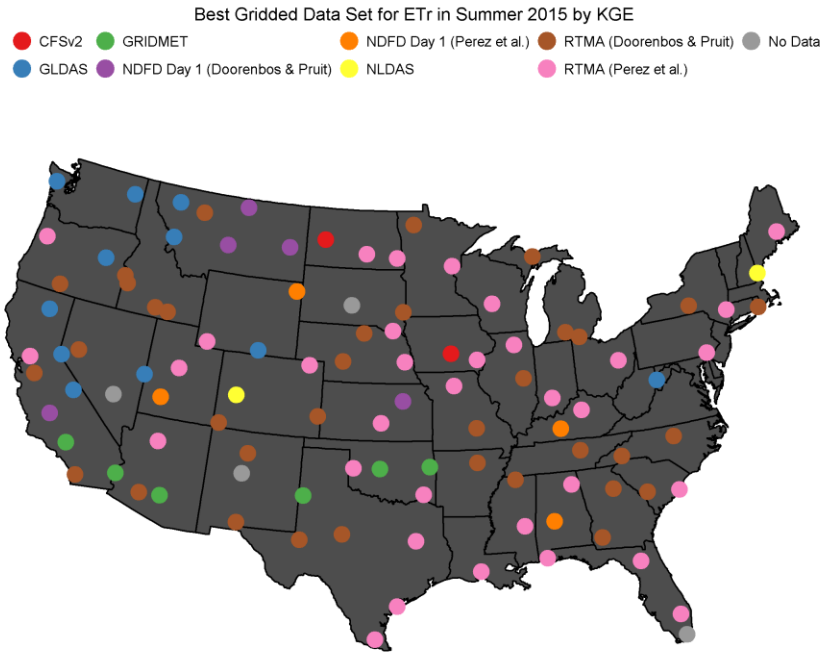

Figure 5.33 Top gridded ET_r data sets ranked by KGE computed using daily data from summer 2015.

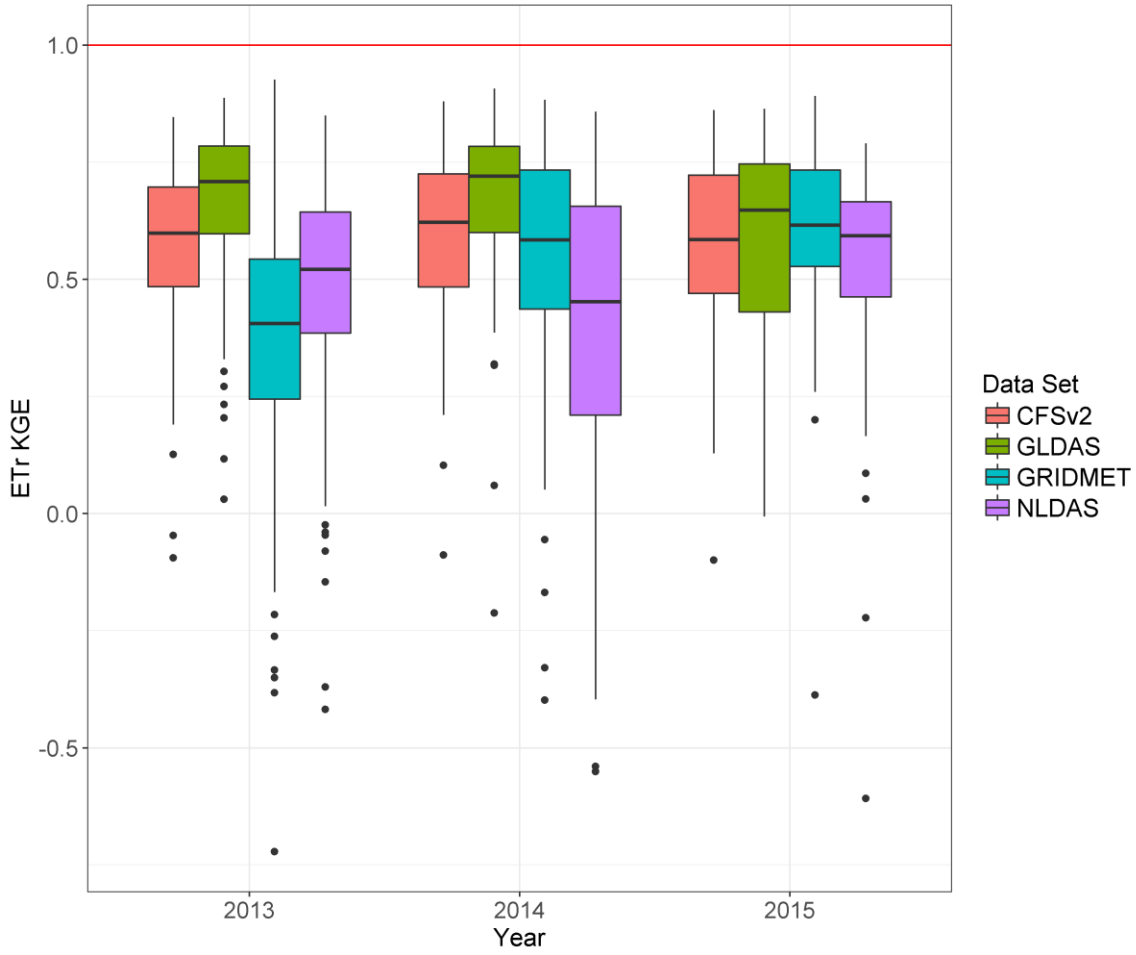

Figure 5.34 ET_r KGE computed for the summer of each study year.

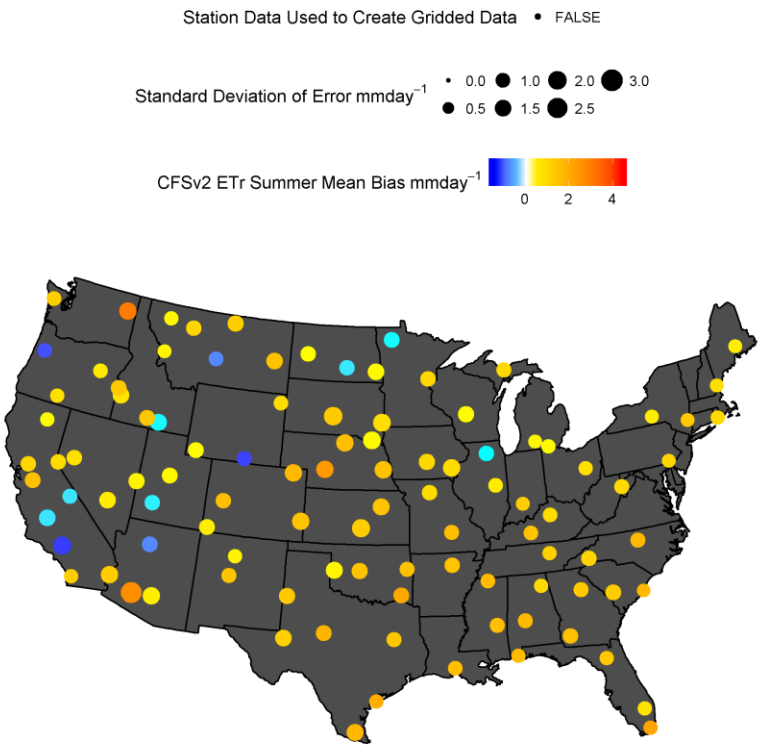

Figure 5.35 CFSv2 ET_r mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

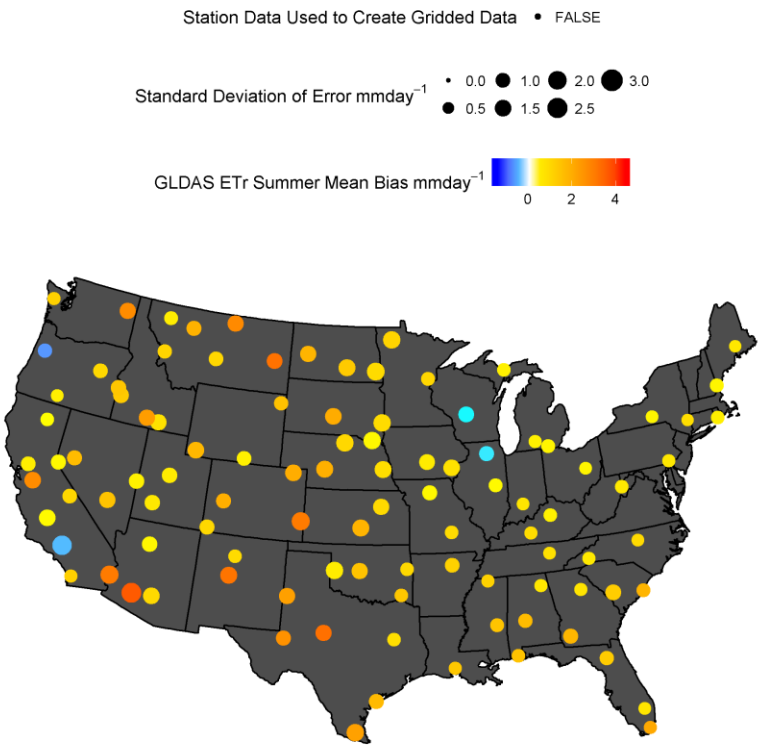

Figure 5.36 GLDAS ET_r mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

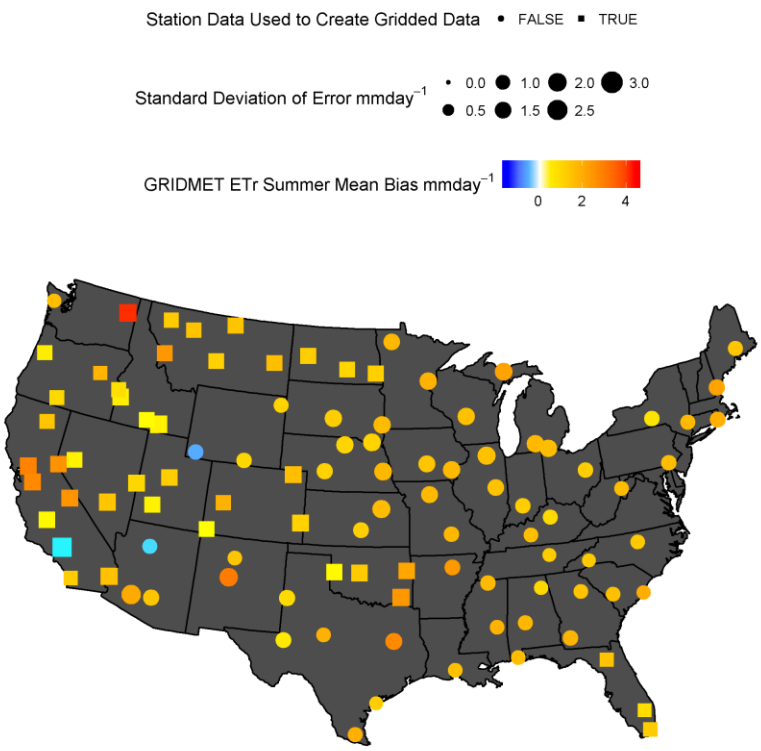

Figure 5.37 GRIDMET ET_r mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

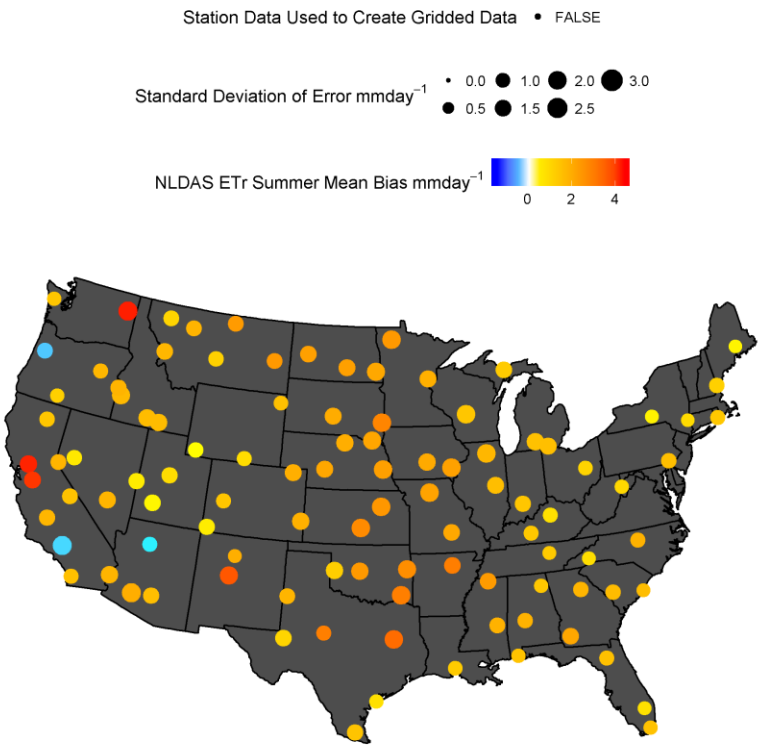

Figure 5.38 NLDAS ET_r mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

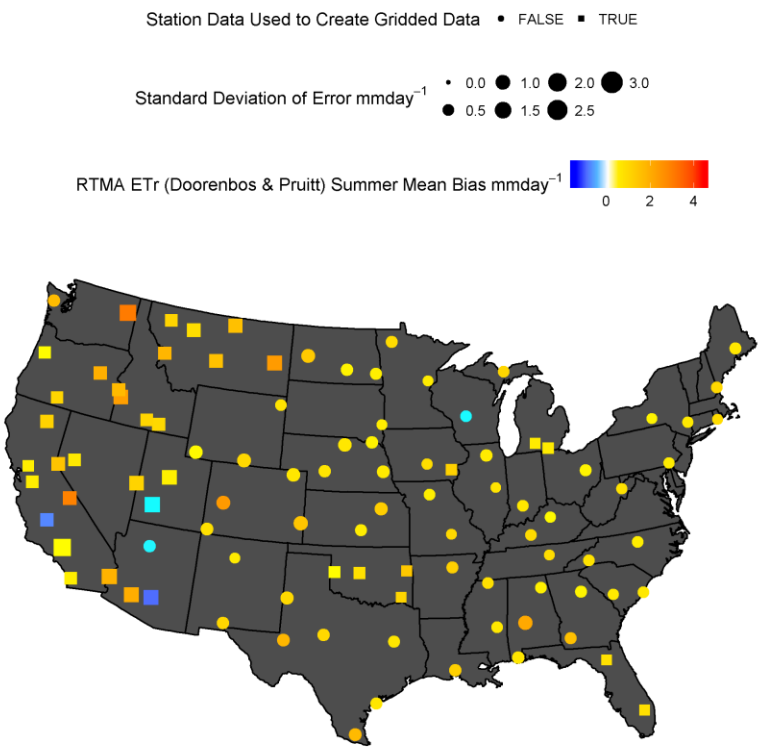

Figure 5.39 RTMA (with radiation estimated with Doorenbos & Pruitt) ET_r mean bias and standard deviation of error computed using summertime daily data for 2015.

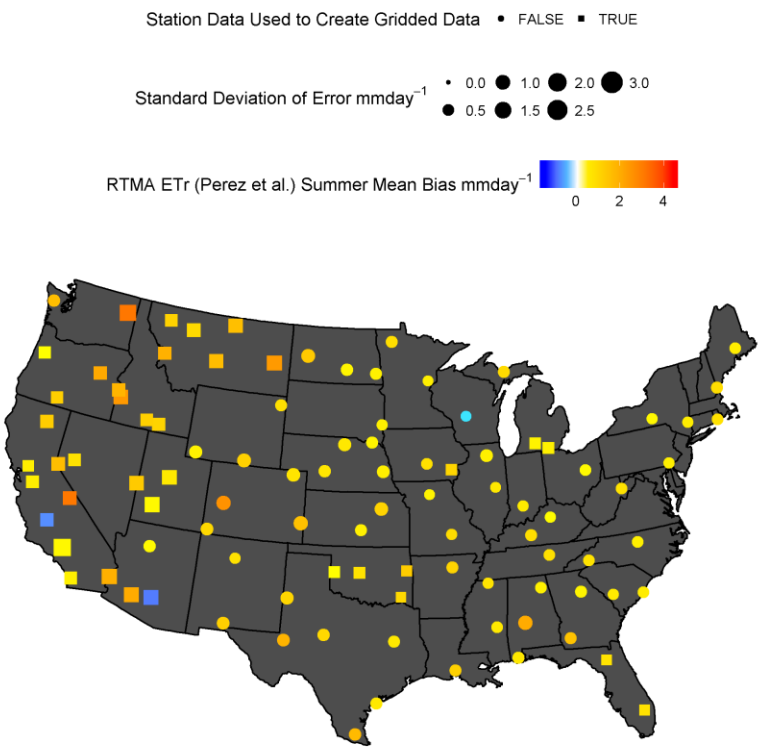

Figure 5.40 RTMA (with radiation estimated with Perez et al.) ET_r mean bias and standard deviation of error computed using summertime daily data for 2015.

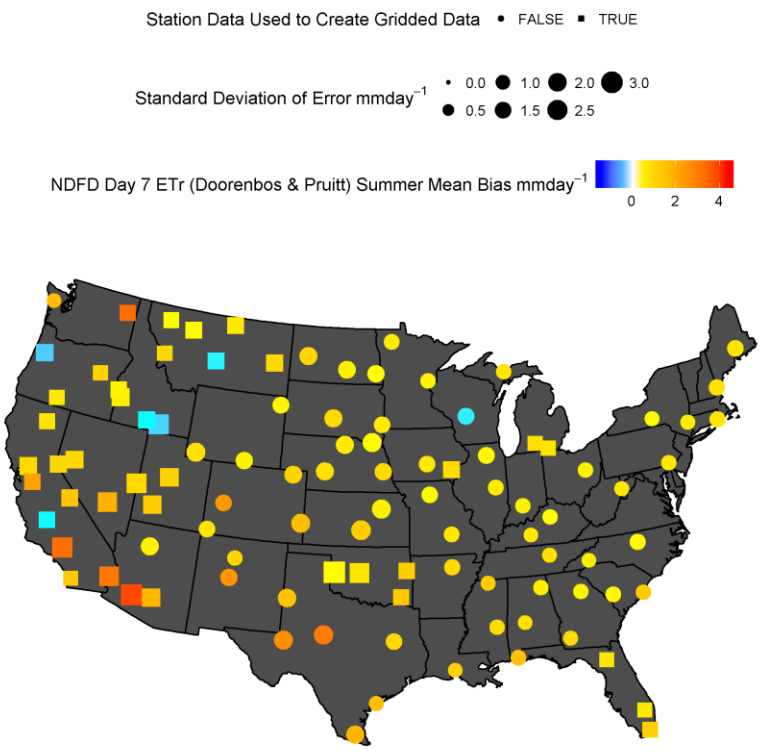

Figure 5.41 NDFD day 7 ET_r (with radiation estimated by Doorenbos & Pruitt) mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

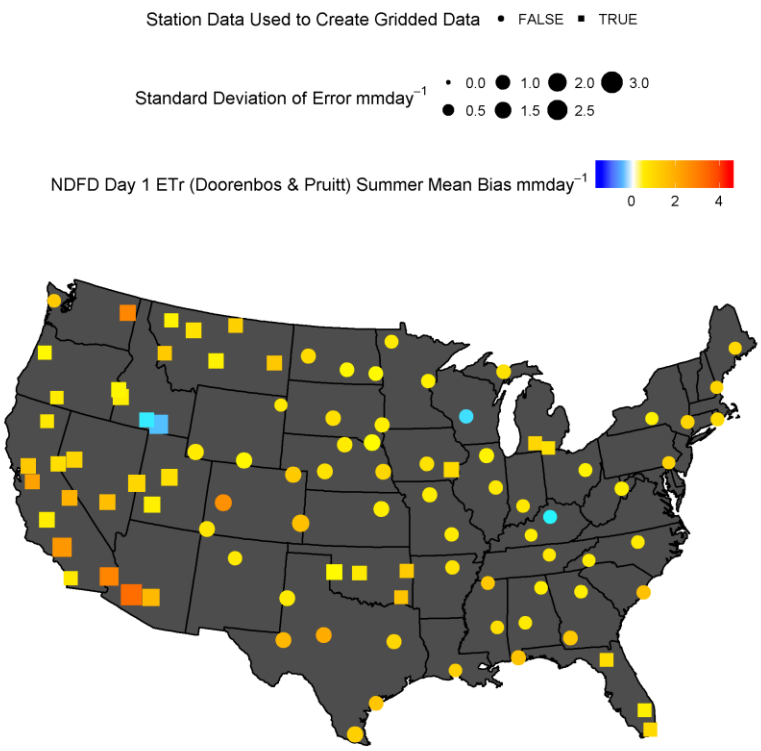

Figure 5.42 NDFD day 1 ET_r (with radiation estimated by Doorenbos & Pruitt) mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

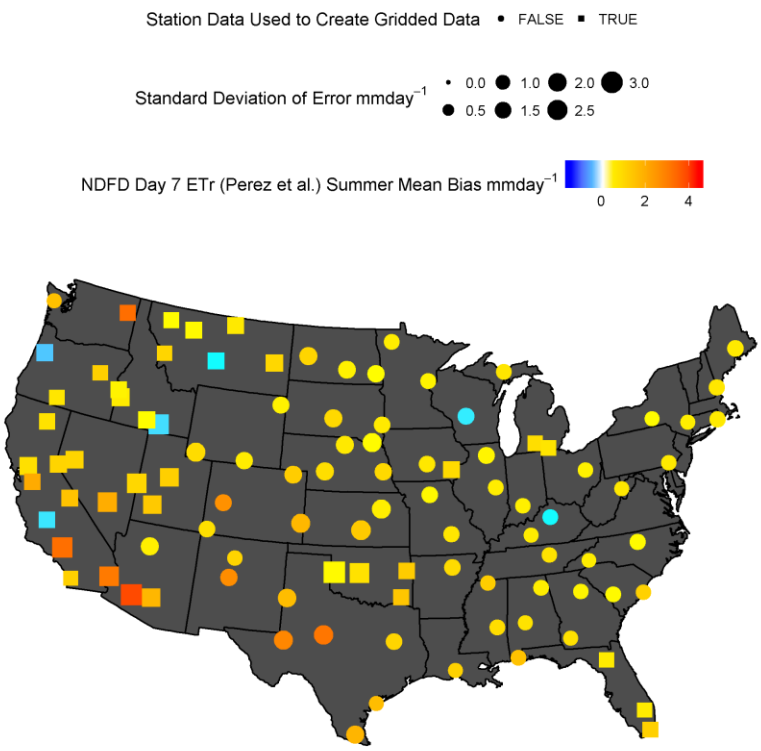

Figure 5.43 NDFD day 7 ET_r (with radiation estimated by Perez et al.) mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

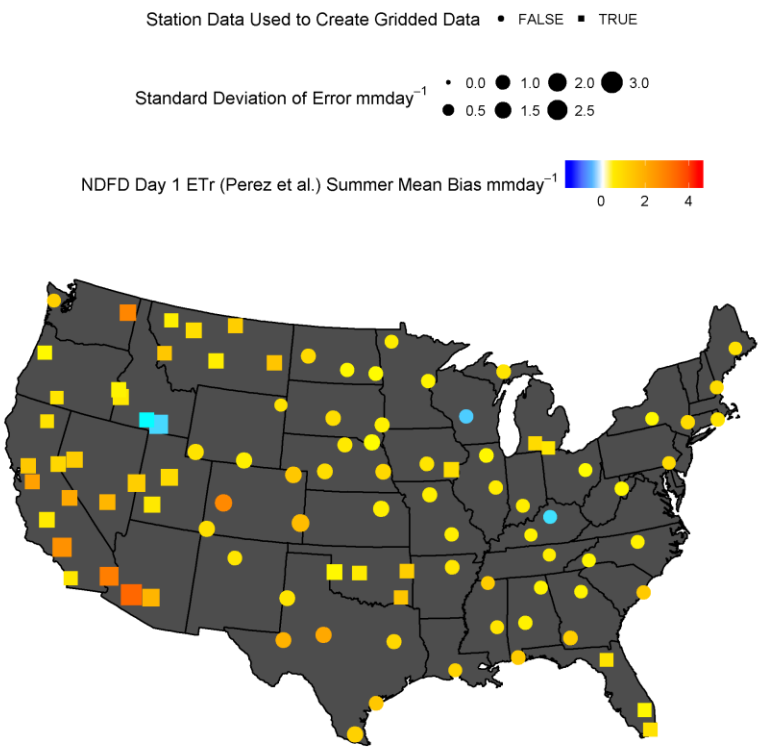

Figure 5.44 NDFD day 1 ET_r (with radiation estimated by Perez et al.) mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

CHAPTER 6. CONCLUSIONS

6.1 Summary

This research evaluated the abilities of several gridded weather data products to estimate ET_{ref} by comparing ET_{ref} computed with those data sets to ET_{ref} computed with ground weather station measurements, where the weather stations were selected at locations expected to be close to reference conditions. Daily vapor pressure, near-surface air temperature, shortwave solar radiation and wind speed - the variables used to compute ET_{ref} - were also compared among the gridded and measured data sets. The results focused on the summer season defined as June, July and August, because this is the primary season during which ET_{ref} is most useful.

Every gridded data set had a tendency to overestimate ET_r at most weather station locations. This is the result of a tendency to overestimate wind speed, solar radiation and temperature and to underestimate vapor pressure as compared to the selected weather stations. RTMA generally had the best KGE - which is a statistic combining mean bias, ratio of standard deviations, and the Pearson correlation - for all weather variables and ET_r. We attribute this good performance to the large amount of surface weather data assimilated into RTMA and the fact that RTMA is an operational analysis and therefore subject to continual improvement.

RTMA started publishing the variable that we used to derive solar radiation, TCDC (total cloud cover %), in April 2015. Before that time, RTMA published cloud cover estimates that were based on the Geostationary Operation Environmental Satellite System (GOES). Deriving solar radiation from those earlier values was not tested in this

thesis. Coupled with the knowledge that the RTMA assimilation system is under continual improvement, and we conclude that RTMA is best used for real-time applications or for applications not aimed at examining long term trends.

NDFD is often initialized and verified using RTMA. NDFD-based ET_r bias was similar to the RTMA bias, possibly as a result of their relationship. The bias of NDFD remained steady between the 1 and 7 day lead times for ET_r, but the correlations dipped from a median of 0.77 (using the Perez et al. 2007 method to convert TCDC to solar radiation) at the 1-day lead time to 0.37 at the 7-day lead time. The variability error shifted from underestimating and overestimating variability to predominantly underestimating variability for the 7-day lead time. In short, there was large ET_r uncertainty at the 7-day lead time.

Among the analysis data sets, the accuracy of ET_r estimated by GLDAS was second best to RTMA. This is attributed to GLDAS ingesting data from the GDAS which, like RTMA, is an operational analysis. Third place was essentially a tie between CFSv2 and GRIDMET. GRIDMET weather variables performed very similar to NLDAS weather variables except for air temperature, which was more accurately estimated due to the bias-correction employed in GRIDMET where NLDAS data are adjusted using air temperature data from PRISM. Except for the summer of 2013, GRIDMET had more accurate temperature estimates than NLDAS, CFSv2 and GLDAS. The good performance of estimated air temperature is what allowed GRIDMET to be equivalent to CFSv2 in the ET_r performance. The CFSv2 is similar to NLDAS because it uses a frozen assimilation system, but the assimilation system used in CFSv2 is more modern than the

one used in NLDAS which helps explain why it performed better than NLDAS despite having greater spatial and temporal resolutions. NLDAS produced the most biased estimates of ET_r likely because it derives its data from the continuation of a 2003 reanalysis (NARR/R-CDAS).

NDFD and RTMA solar radiation had reduced variability error when the conversion from sky cover to radiation was made using the equation derived in Perez et al. (2007) instead of Doorenbos & Pruitt (1977). No benefits of using the Perez et al. equation were seen in the ET_r estimates. To the contrary, the ET_r variability error increased when using Perez et al. instead of Doorenbos & Pruitt. The underestimation of radiation variability by Doorenbos & Pruitt likely counteracted an overestimation of variability in another variable, such as wind speed. In order for ET_r estimates to realize any benefit from utilizing the Perez et al. equation the variability error in other variables need to be reduced.

6.2 Conclusions

There are several potential explanations for the biases observed in the gridded data sets. One is that gridded data sets primarily assimilate ground based weather station data from airports which may have hotter and less humid near-surface air properties than agricultural areas in the surrounding region. Another is the inability of gridded data to incorporate irrigation into their soil water balances, which results in an underestimation of ET from irrigated areas and its cooling and humidifying effects on the near-surface boundary layer. Lastly, there may be biases within the models themselves due to incorrect parameterizations, unrepresented physics or inaccurate boundary forcing

(Menard, 2010). It is important to note that the gridded weather data sets may show less bias when compared to airport and other data sets that may be more representative of the overall region and climate. Our study and comparisons with ground-based weather data focused on weather data collected from locations that were likely to exhibit the cooling and humidifying characteristics associated with well-watered vegetation. This focus was intentional because ET_{ref} equations such as the ASCE Penman-Monteith method have been developed to 'expect' weather data that reflect conditions above a reference or near-reference surface. Application of ET_{ref} equations to weather data collected over dry surfaces causes overstatement of the actual ET_{ref}, which by definition, occurs over a reference surface (well-watered, full-cover vegetation) (Walter et al., 2005; Jensen & Allen, 2015).

None of the six gridded weather data sets produced calculations of ET_r that were equivalent to ET_r calculated from observed weather data. The gridded weather data sets all produced median ET_r biases ranging from positive 10 to positive 25%. Some locations may have consistently low bias and could be used without any form of bias correction, but generally, the gridded data are in need of bias correction to be useful and accurate.

6.3 Recommendations

Future improvements to data assimilation systems will no doubt reduce the error identified in this study. Evidence of past improvements are manifested in GLDAS's and CFSv2's superior performance to NLDAS despite having coarser spatial and temporal resolutions. At present, the error in the gridded data analyzed in this study is problematic

for the estimation of ET_{ref}. At a minimum, some of the techniques in the methods section of this study should be employed to better match the gridded data to the location at which ET_{ref} needs to be computed. Namely, adjusting the gridded temperature and humidity values from the grid cell elevation to the elevation at the area of interest, carefully handling timestamps and time representations, and converting instantaneous solar radiation values to time averaged values using theoretical clear-sky radiation.

This study adds to the collection of information available regarding gridded weather data sets and their error, but there are additional methods to explore for error and other gridded data sets to examine. Lewis et al. used spatial bilinear interpolation of the gridded data to the weather station locations to better match the spatial scale of the weather station data. This study may have benefited from this procedure. Future studies could quantify any reductions in the differences between gridded and station data resulting from spatial interpolation.

Along a similar line, Diremeyer et al. (2016) investigated how representative point measurements of soil moisture were of a gridded product by locating grid cells containing many soil moisture instruments. They averaged the point measurements together one by one, and observed how quickly the mean converged to a stable value. Averaging only a small number of point measurements before coming to a stable value near the grid cell value indicated wide spatial representivity. This kind of analysis would be beneficial to apply to gridded weather data because it measures the error arising from mismatching spatial scales. However, it is relatively rare to find multiple weather stations within a single grid cell.

Another improvement to comparisons between gridded data and observed weather data may be the use of a spline function to do time interpolation between the observed data to simulate the instantaneous values in gridded weather data. Since most weather variables studied tend to have an upward convex trend during daytime, and downward convex trend during nighttime, the linear interpolation used in this thesis tends to underestimate instantaneous values during daytime and overestimate instantaneous values during nighttime. A spline function may follow expected change in weather values better, especially for near-surface air temperature.

Other future research could pursue the important task of debiasing gridded weather data or debiasing ET_{ref} directly. This might include developing statistical models for bias or using more theoretical adjustments based on surface to air profiles for temperature, humidity and wind as described by Allen, Kilic, DeBruin and Joros (2013). Debiasing gridded data would allow the data to be used to its fullest potential, and aid modeling activities that benefit from spatially and temporally complete coverage.

CHAPTER 7. REFERENCES

References

Abatzoglou, J. T. (2011). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology, 33*(1), 121-131. doi:10.1002/joc.3413

Allen, R. G. (2016). In Li G. Z. (Ed.), *Ref-et* (4.1st ed.) University of Idaho.

Allen, R. G., Kilic, A., DeBruin, H., & Joros, A. (2013). Conditioning of NLDAS, NARR and arid weather station data to improve their representation of well-watered (reference) environments associated with irrigated agriculture. *93rd American Meteorological Society Annual Meeting,*

Allen, R. G., Tasumi, M., & Trezza, R. (2007). Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - model. *Journal of Irrigation and Drainage Engineering, 133*(4), 380-394. doi:10.1061/(ASCE)0733-9437(2007)133:4(380)

Allen, R. G. (1996). Assessing integrity of weather data for reference evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering-Asce, 122*(2), 97-106. doi:10.1061/(ASCE)0733-9437(1996)122:2(97)

Bastiaanssen, W., Menenti, M., Feddes, R., & Holtslag, A. (1998). A remote sensing surface energy balance algorithm for land (SEBAL) - 1. formulation. *Journal of Hydrology, 212*(1-4), 198-212. doi:10.1016/S0022-1694(98)00253-4

Benjamin, S., Devenyi, D., Weygandt, S., Brundage, K., Brown, J., Grell, G., . . . Manikin, G. (2004). An hourly assimilation-forecast cycle: The RUC. *Monthly Weather Review, 132*(2), 495-518. doi:10.1175/1520-0493(2004)1322.0.CO;2

Bird, R. E., & Hulstrom, R. L. (1981). *A simplified clear sky model for direct and diffuse insolation on horizontal surfaces.* (Technical Report No. TR-642-761). Golden, Colorado: Solar Energy Research Institute.

Charney, J. J., Zhong, S., Kiefer, M. T., & Zhu, X. (2013). *An investigation of the differences between real time mesoscale analysis and observed meteorological conditions at RAWS stations in the northeast united states.* Unpublished manuscript.

Cosgrove, B., Lohmann, D., Mitchell, K., Houser, P., Wood, E., Schaake, J., . . . Meng, J. (2003). Real-time and retrospective forcing in the north american land data assimilation system (NLDAS) project. *Journal of Geophysical Research-Atmospheres, 108*(D22), 8842. doi:10.1029/2002JD003118

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., . . . Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. *International Journal of Climatology, 28*(15), 2031-2064. doi:10.1002/joc.1688

De Pondeca, M. S. F. V., Manikin, G. S., DiMego, G., Benjamin, S. G., Parrish, D. F., Purser, R. J., . . . Vavra, J. (2011). The real-time mesoscale analysis at NOAA's national centers for environmental prediction: Current status and development. *Weather and Forecasting, 26*(5), 593-612. doi:10.1175/WAF-D-10-05037.1

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., & Bosilovich, M. G. (2012). Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal of Climate, 25*(6), 1916-1944. doi:10.1175/JCLI-D-11-00004.1

Dirmeyer, P. A., Wu, J., Norton, H. E., Dorigo, W. A., Quiring, S. M., Ford, T. W., . . . Lawrence, D. M. (2016). Confronting weather and climate models with observational data from soil moisture networks over the united states. *Journal of Hydrometeorology, 17*(4), 1049-1067. doi:10.1175/JHM-D-15-0196.1

Doorenbos, J., & Pruitt, W. O. (1977). Crop water requirements. *FAO Irrigation and Drainage Paper, 24*, 8-15.

Glahn, H., & Ruth, D. (2003). The new digital forecast database of the national weather service. *Bulletin of the American Meteorological Society, 84*(2), 195-201. doi:10.1175/BAMS-84-2-195

Google Earth Engine Team. (2015). Google earth engine: A planetary-scale geo-spatial analysis platform. Retrieved from https://earthengine.google.com

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology, 377*(1-2), 80-91. doi:10.1016/j.jhydrol.2009.08.003

Hobbins, M. T. (2016). The variability of asce standardized reference evapotranspiration: A rigorous, conus-wide decomposition and attribution. *Transactions of the Asabe, 59*(2), 561-576.

Houser, P. R., Lannoy, G. L. M., & Walker, J. P. (2010). Land surface data assimilation. In W. Lahoz, B. Khattatov & R. Menard (Eds.), *Data assimilation: Making sense of observations* (pp. 549) Springer. doi:10.1007/978-3-540-74703-1

Hupet, F., & Vanclooster, M. (2001). Effect of the sampling frequency of meteorological variables on the estimation of the reference evapotranspiration. *Journal of Hydrology, 243*(3-4), 192-204. doi:10.1016/S0022-1694(00)00413-3

Irmak, A., Allen, R. G., Kjaersgaard, J., Huntington, J., Kamble, B., Trezza, R., & Ratcliffe, I. (2011). Operational remote sensing of ET and challenges. In A. Irmak (Ed.), *Evapotranspiration - remote sensing and modeling* (pp. 467). Rijeka, Croatia: InTech.

Jensen, M. E., & Allen, R. G. (Eds.). (2015). *Evaporation, evapotranspiration, and irrigation water requirements. ASCE manuals and reports on engineering practice no. 70.* (2nd ed.). Reston, Virginia: American Society of Civil Engineers.

Kilic, A., Huntington, J., Allen, R. G., Morton, C., Kamble, B., Gorelick, N., . . . Robinson, C. (2014). An evapotranspiration mapping tool at landsat resolution on the google earth engine: EEFlux. *An Evapotranspiration Mapping Tool at Landsat Resolution on the Google Earth Engine: EEFlux,*

Lahoz, W., Khattatov, B., & Menard, R. (2010). Data assimilation and information. In W. Lahoz, B. Khattatov & R. Menard (Eds.), *Data assimilation: Making sense of observations* (pp. 3) Springer. doi:10.1007/978-3-540-74703-1

Lewis, C. S., Geli, H. M. E., & Neale, C. M. U. (2014). Comparison of the NLDAS weather forcing model to agrometeorological measurements in the western united states. *Journal of Hydrology, 510,* 385-392. doi:10.1016/j.jhydrol.2013.12.040

Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research, 50*(2), 1131-1151. doi:10.1002/2013WR014581

McEvoy, D. J., Mejia, J. F., & Huntington, J. L. (2014). Use of an observation network in the great basin to evaluate gridded climate data. *Journal of Hydrometeorology, 15*(5), 1913-1931. doi:10.1175/JHM-D-14-0015.1

Menard, R. (2010). Bias estimation. In R. Menard, W. Lahoz & B. Khattatov (Eds.), *Data assimilation: Making sense of observations* (pp. 113) Springer. doi:10.1007/978-3-540-74703-1_6

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P., Ebisuzaki, W., . . . Shi, W. (2006). North american regional reanalysis. *Bulletin of the American Meteorological Society, 87*(3), 343-360. doi:10.1175/BAMS-87-3-343

Mitchell, K., Marshall, C., Lohmann, D., Ek, M., Lin, Y., Grunmann, P., . . . Duan, Q. (2000). *The collaborative GCIP land data assimilation (LDAS) project and supportive NCEP uncoupled land-surface modeling initiatives*. BOSTON; 45 BEACON ST, BOSTON, MA 02108 USA: AMER METEOROLOGICAL SOCIETY.

Moorhead, J., Gowda, P., Hobbins, M., Senay, G., Paul, G., Marek, T., & Porter, D. (2015). Accuracy assessment of NOAA gridded daily reference evapotranspiration for the texas high plains. *Journal of the American Water Resources Association, 51*(5), 1262-1271. doi:10.1111/1752-1688.12303

Myrick, D. T., & Horel, J. D. (2006). Verification of surface temperature forecasts from the national digital forecast database over the western united states. *Weather and Forecasting, 21*(5), 869-892. doi:10.1175/WAF946.1

Palmer, C., Osborne, H., & Hobbins, M. T. (2017). Using FRET for drought mitigation. *American Meteorological Society Annual Meeting*, Seattle, WA.

Pearce, S. (2016a). Data sources summary. Retrieved from http://ldas.gsfc.nasa.gov/gldas/GLDASforcing.php

Pearce, S. (2016b). GLDAS: Project goals. Retrieved from http://ldas.gsfc.nasa.gov/gldas/GLDASgoals.php

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Jr., Renne, D., & Hoff, T. E. (2010). Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy, 84*(12), 2161-2172. doi:10.1016/j.solener.2010.08.014

Perez, R., Moore, K., Wilcox, S., Renne, D., & Zelenka, A. (2007). Forecasting solar radiation - preliminary evaluation of an approach based upon the national forecast database. *Solar Energy, 81*(6), 809-812. doi:10.1016/j.solener.2006.09.009

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., . . . Toll, D. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society, 85*(3), 381-+. doi:10.1175/BAMS-85-3-381

Rood, R. (2010). The role of the model in the data assimilation system. In W. Lahoz, B. Khattatov & R. Menard (Eds.), *Data assimilation: Making sense of observations* (pp. 351) Springer. doi:10.1007/978-3-540-74703-1

Saha, S., Moorthi, S., Pan, H., Wu, X., Wang, J., Nadiga, S., . . . Goldberg, M. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society, 91*(8), 1015-1057. doi:10.1175/2010BAMS3001.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., . . . Becker, E. (2014). The NCEP climate forecast system version 2. *Journal of Climate, 27*(6), 2185-2208. doi:10.1175/JCLI-D-12-00823.1

Slater, A. G. (2016). Surface solar radiation in north america: A comparison of observations, reanalyses, satellite, and derived products. *Journal of Hydrometeorology, 17*(1), 401-420. doi:10.1175/JHM-D-15-0087.1

Walter, I. A., Allen, R. G., Elliott, R., Itenfisu, D., Brown, P., Jensen, M. E., . . . Wright, J. L. (2005). The ASCE standardized reference evapotranspiration equation. *ASCE-EWRI Task Committee Report,*

Wang, K., & Dickinson, R. E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics, 50*, RG2005. doi:10.1029/2011RG000373

CHAPTER 8. APPENDICES

Appendix A. Data Dictionary

ADAS	Advanced Regional Prediction System Data Assimilation System
AFWA	Air Force Weather Agency
CDASv2	Climate Data Assimilation System version 2
CFSR	Climate Forecast System Reanalysis
CFSv2	Climate Forecast System version 2
CLM	Community Land Model
CMAP	CPC Merged Analysis of Precipitation
CONUS	Continental United States
CPC	Climate Prediction Center
CSR	Clear Sky Ratio
CST	Central Standard Time
EEFlux	Earth Engine Flux (Implementation of METRIC)
ERC	Energy Release Component
ESRL	Earth System Research Laboratory
ET	Evapotranspiration
ET_o	Grass Reference Evapotranspiration
ET_r	Alfalfa Reference Evapotranspiration
ET_{ref}	Reference Evapotranspiration
ET_{sz}	Standardized Penman-Monteith Reference Evapotranspiration
FLUXNET	A network of regional networks of eddy covariance towers
GDAS	Global Data Assimilation System
GFS	Global Forecast System
GHCN	Global Historical Climate Network
GLDAS	Global Land Data Assimilation System
GOES	Geostationary Operational Environmental Satellite
GRIB	Gridded Binary
GRIDMET	University of Idaho Gridded Surface Meteorological Data Set
GSFC	Goddard Space Flight Center
KGE	Kling-Gupta Efficiency
KSI	Kolmogorov-Smirnov test Integral
LDAS	Land Data Assimilation System
LSM	Land Surface Model
MAE	Mean Absolute Error
MBE	Mean Bias Error

METRIC	Mapping Evapotranspiration at high Resolution with Internalized Calibration
MODIS	Moderate Resolution Imaging Spectroradiometer
MST	Mountain Standard Time
NARR	North American Regional Reanalysis
NASA	National Aeronautical and Space Administration
NCEP	National Center for Environmental Prediction
NDFD	National Digital Forecast Database
NDVI	Normalized Difference Vegetation Index
NED	National Elevation Dataset
NESDIS	National Environmental, Satellite, and Data Information Service
NLDAS	North American Land Data Assimilation System
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
NWS	National Weather Service
PRISM	Parameter-elevation Relationships on Independent Slopes Model
PST	Pacific Standard Time
QAQC	Quality Assurance and Quality Control
R_a	Theoretical incoming extraterrestrial shortwave solar radiation
RAWS	Remote Automated Weather Stations
R-CDAS	Regional Climate Data Assimilation System
RH	Relative Humidity
RMSE	Root Mean Squared Error
R_s	Incoming shortwave solar radiation
RS	Remote Sensing
R_{so}	Theoretical incoming clear sky shortwave solar radiation
RTMA	Real-time Mesoscale Analysis
RUC	Rapid Update Cycle
SDE	Standard Deviation of Error
SURFRAD	Surface Radiation network
UTC	Universal Coordinated Time
VIC	Variable Infiltration Capacity model
WFO	Weather Forecast Office

Appendix B. Summer Shortwave Solar Radiation Bias Maps

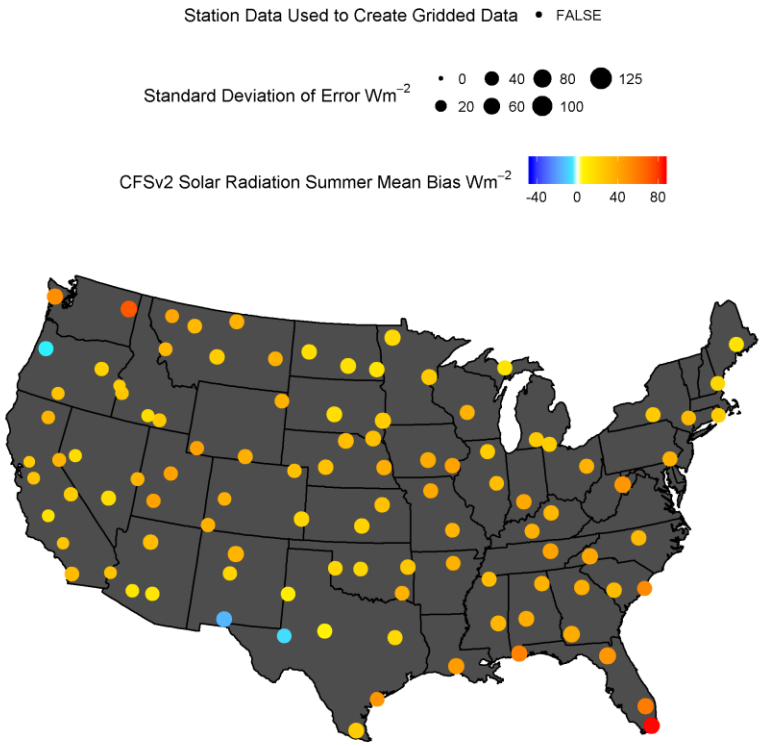

Figure 8.1 CFSv2 shortwave solar radiation mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

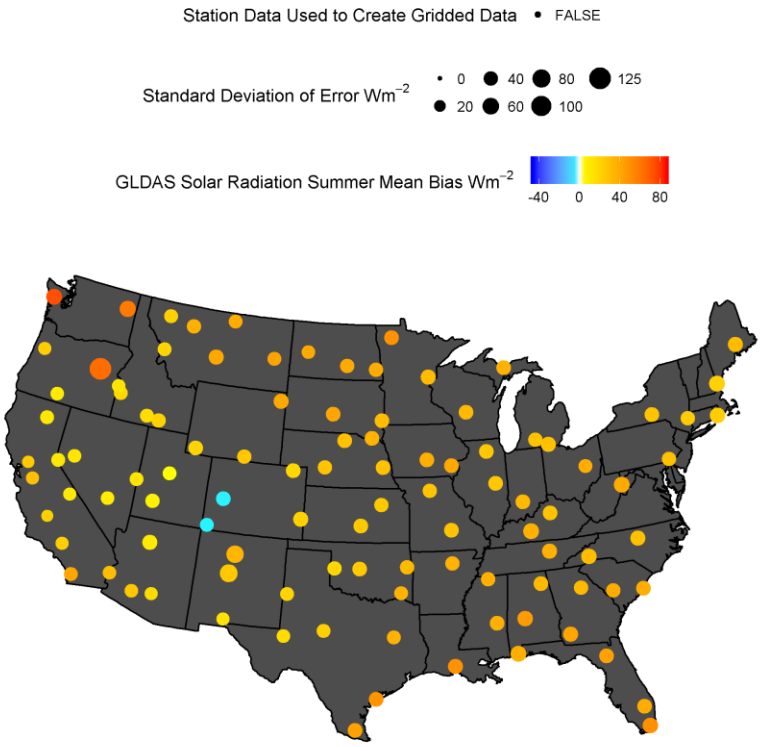

Figure 8.2 GLDAS shortwave solar radiation mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

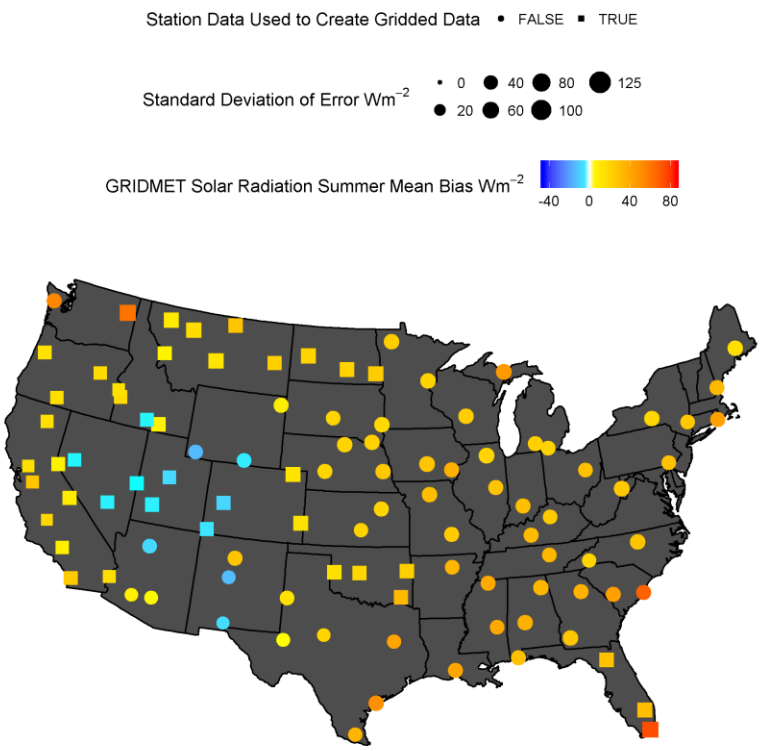

Figure 8.3 GRIDMET shortwave solar radiation mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

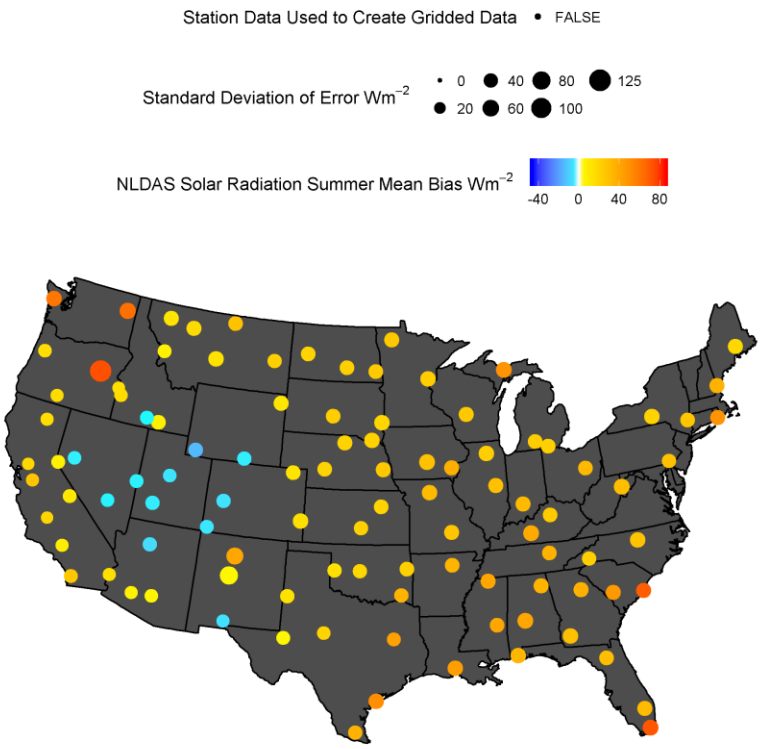

Figure 8.4 NLDAS shortwave solar radiation mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

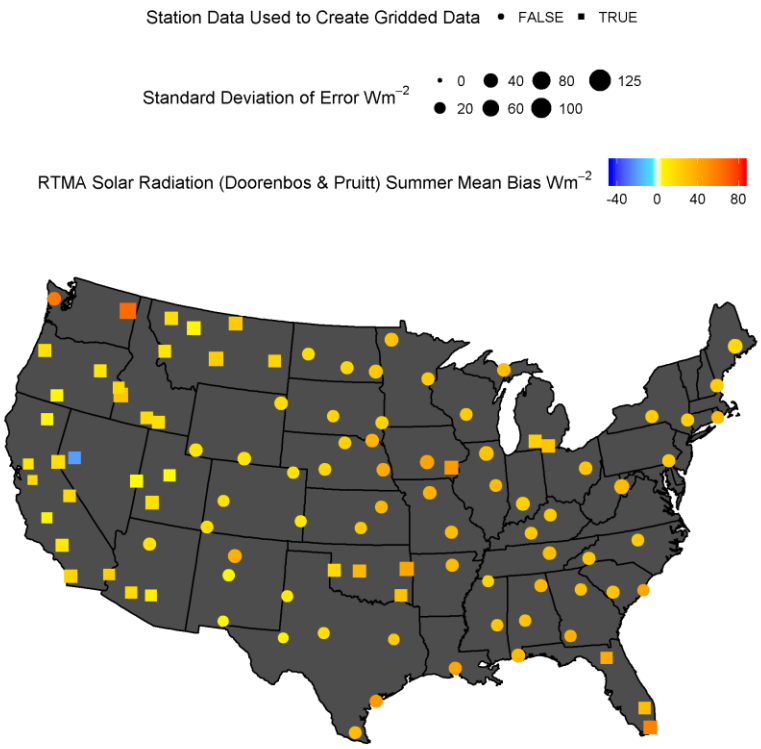

Figure 8.5 RTMA shortwave solar radiation estimated using the Doorenbos & Pruitt method mean bias and standard deviation of error computed using summertime daily data for 2015.

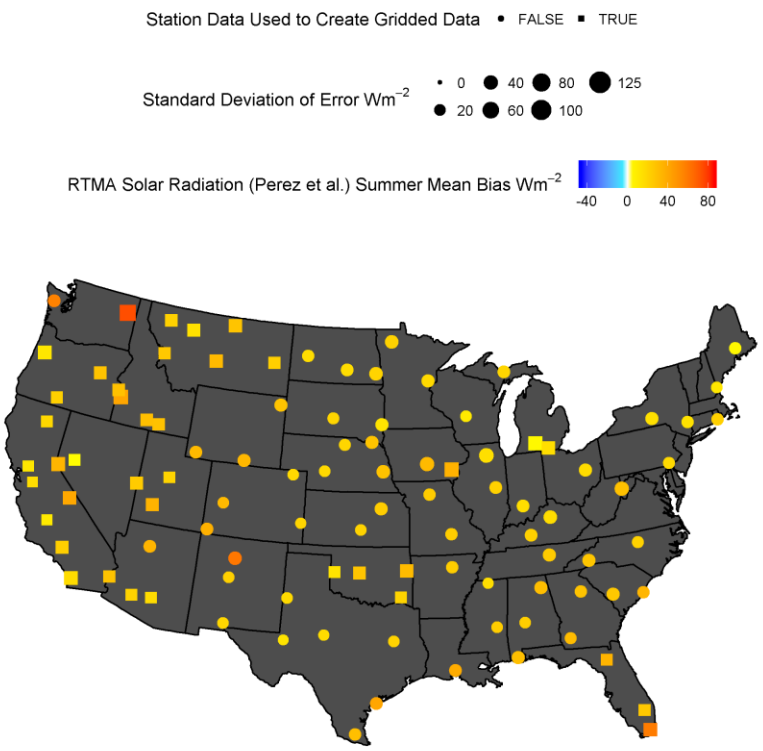

Figure 8.6 RTMA shortwave solar radiation estimated using the Perez et al. method mean bias and standard deviation of error computed using summertime daily data for 2015.

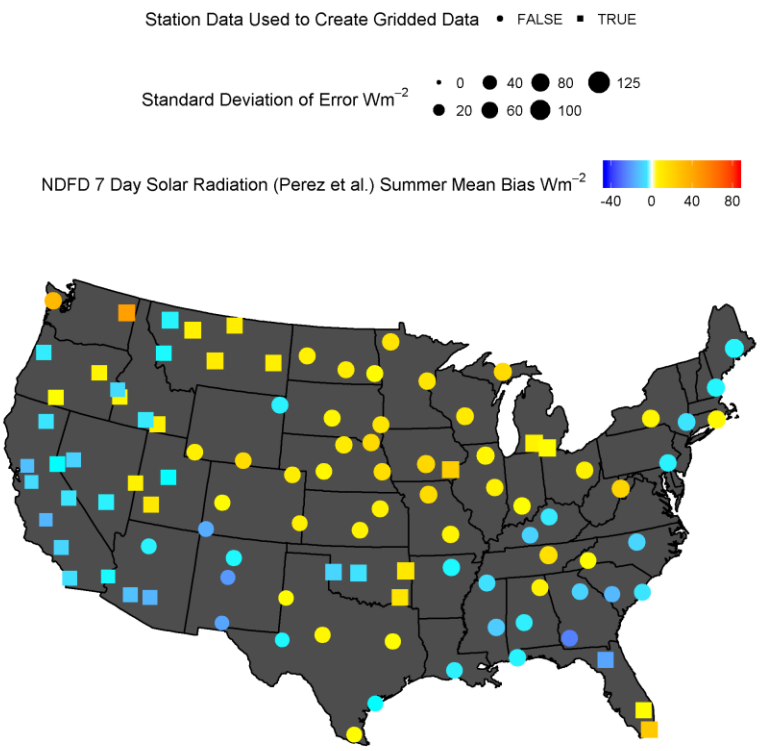

Figure 8.7 NDFD day 7 shortwave solar radiation estimated using the Perez et al. method mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

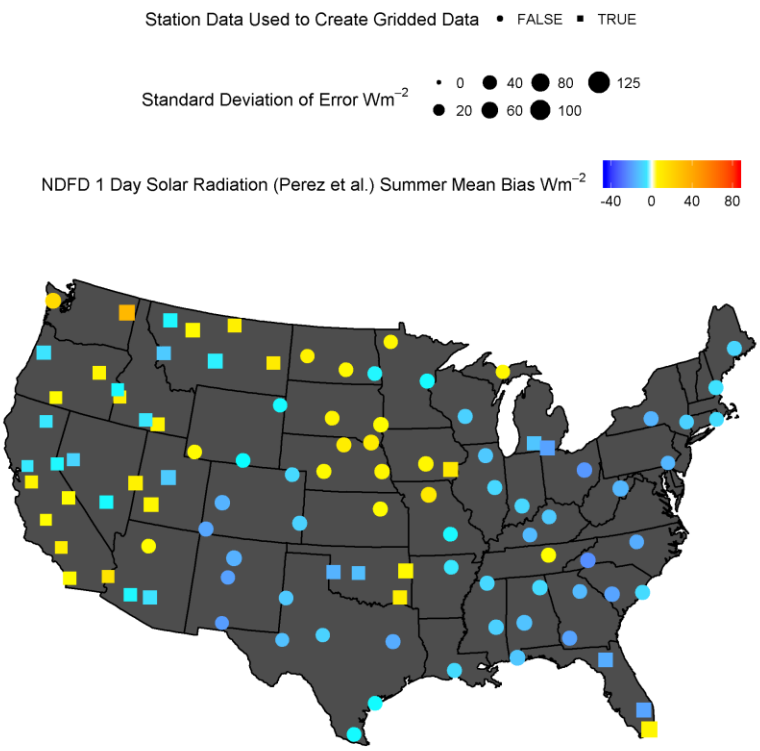

Figure 8.8 NDFD day 1 shortwave solar radiation estimated using the Perez et al. method mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

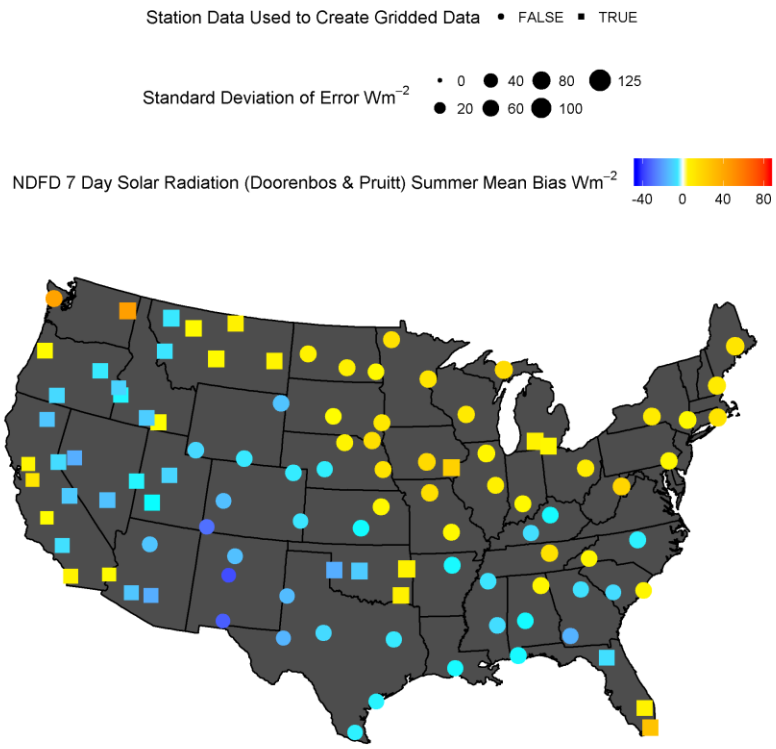

Figure 8.9 NDFD day 7 shortwave solar radiation estimated using the Doorenbos & Pruitt method mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

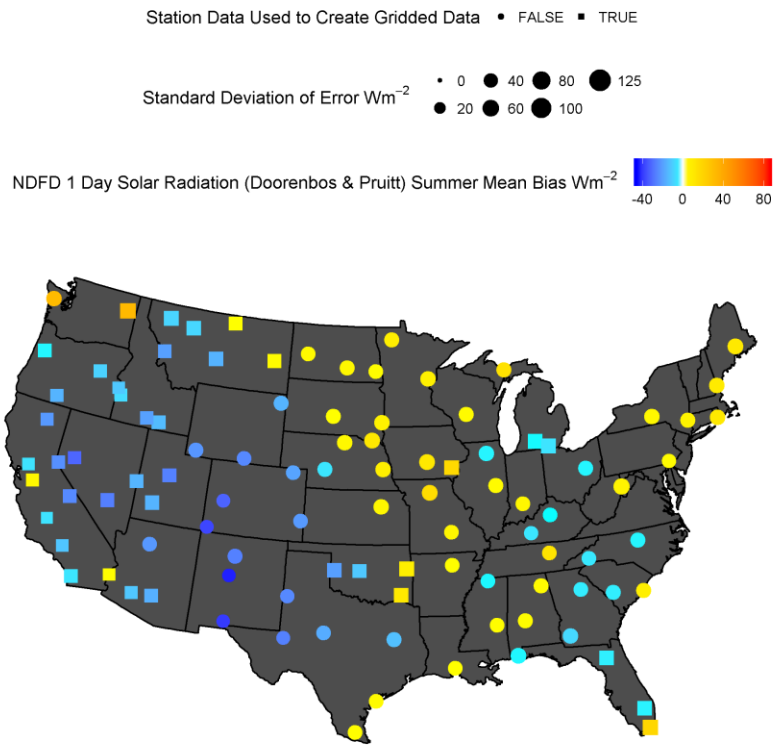

Figure 8.10 NDFD day 1 shortwave solar radiation estimated using the Doorenbos & Pruitt method mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

Appendix C. Summer Wind Speed Bias Maps

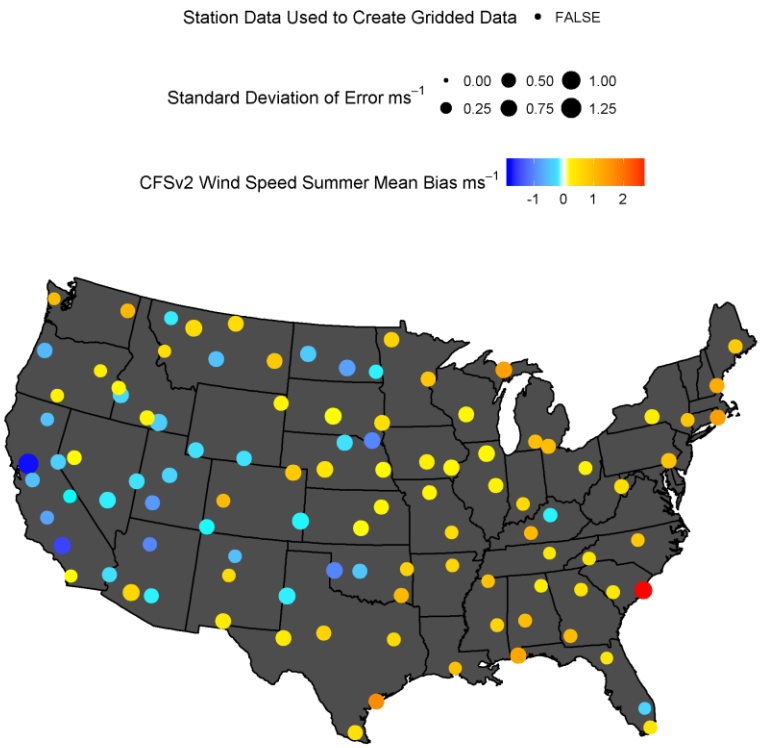

Figure 8.11 CFSv2 wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

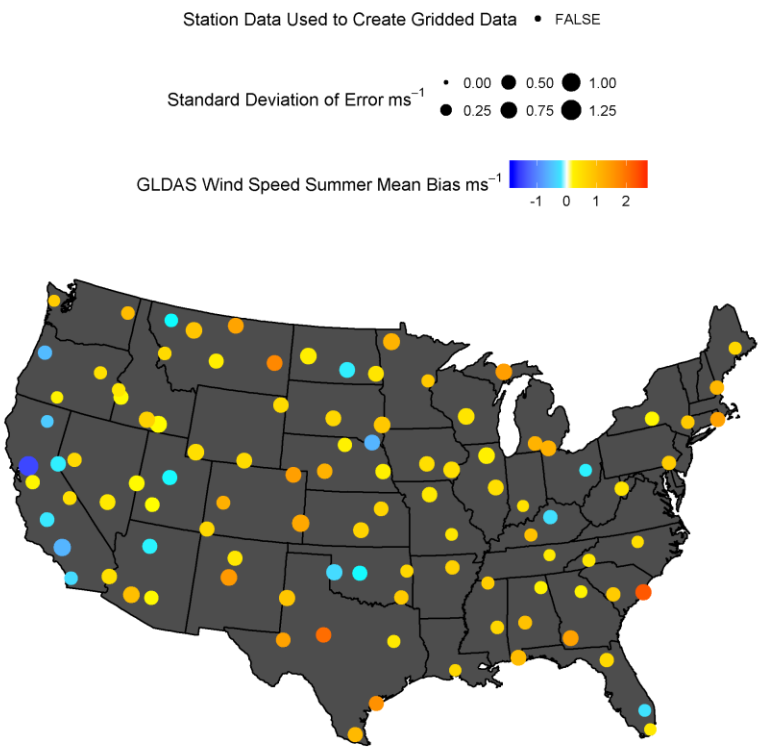

Figure 8.12 GLDAS wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

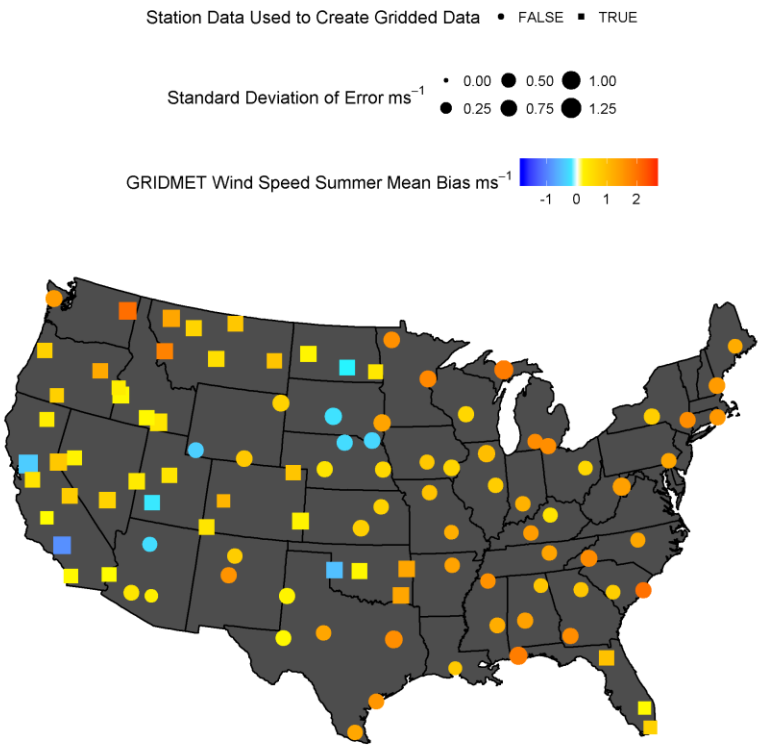

Figure 8.13 GRIDMET wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

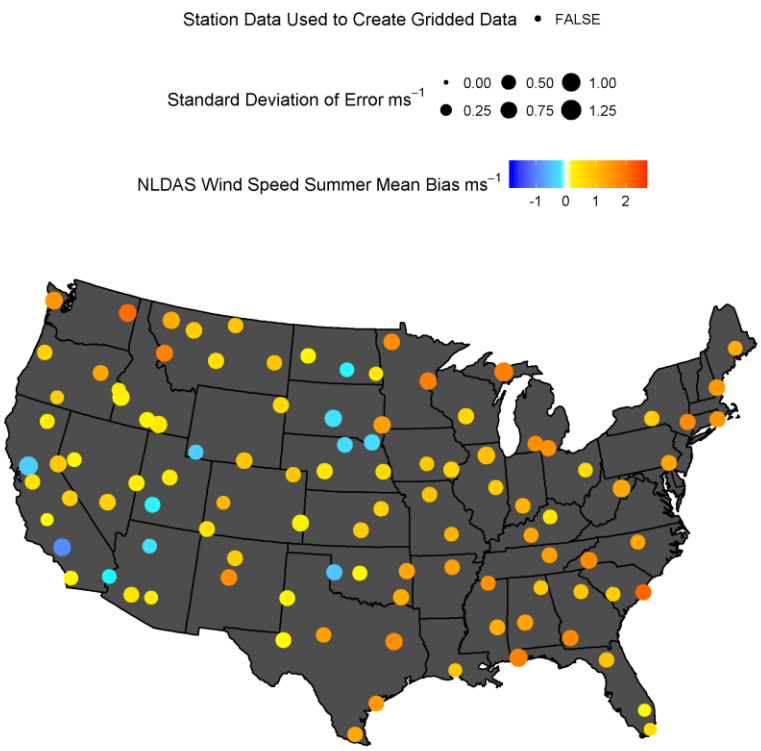

Figure 8.14 NLDAS wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

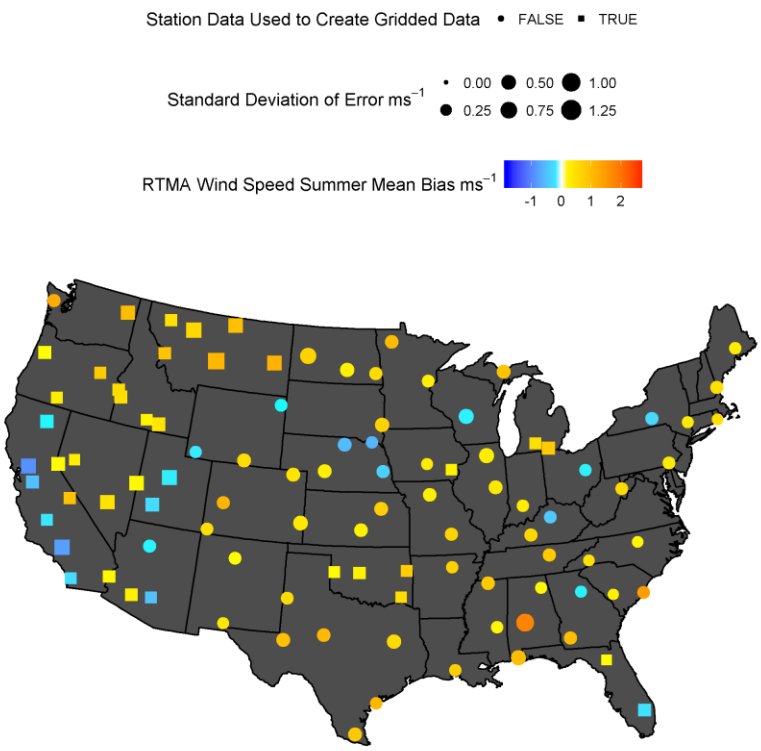

Figure 8.15 RTMA wind speed mean bias and standard deviation of error computed using summertime daily data for 2015.

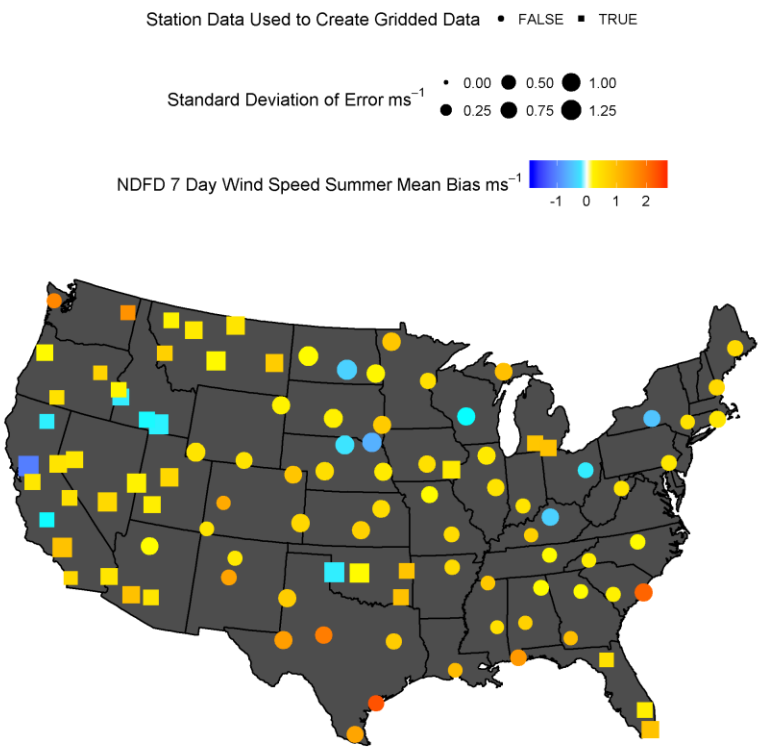

Figure 8.16 NDFD day 7 wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

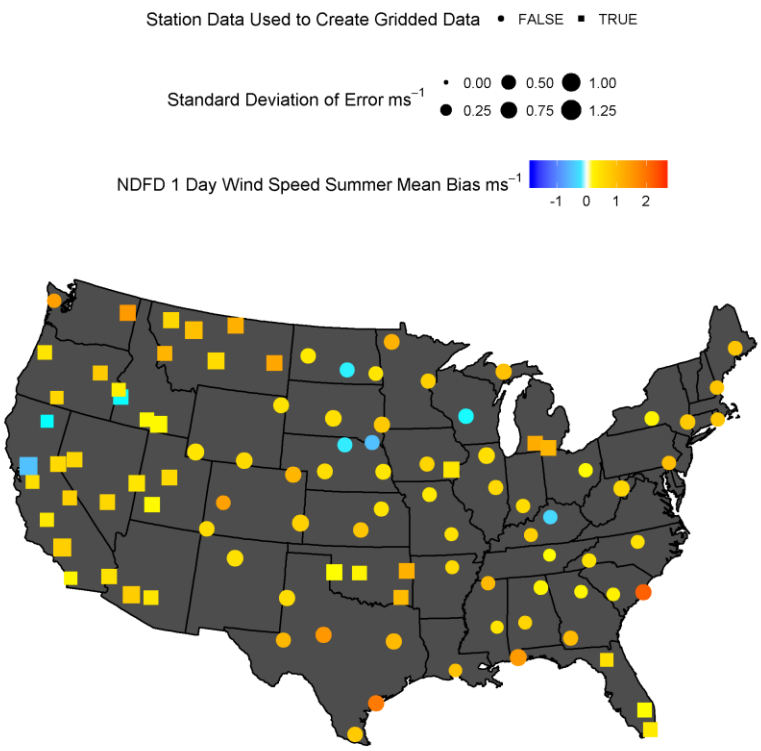

Figure 8.17 NDFD day 1 wind speed mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

Appendix D. Summer Vapor Pressure Bias Maps

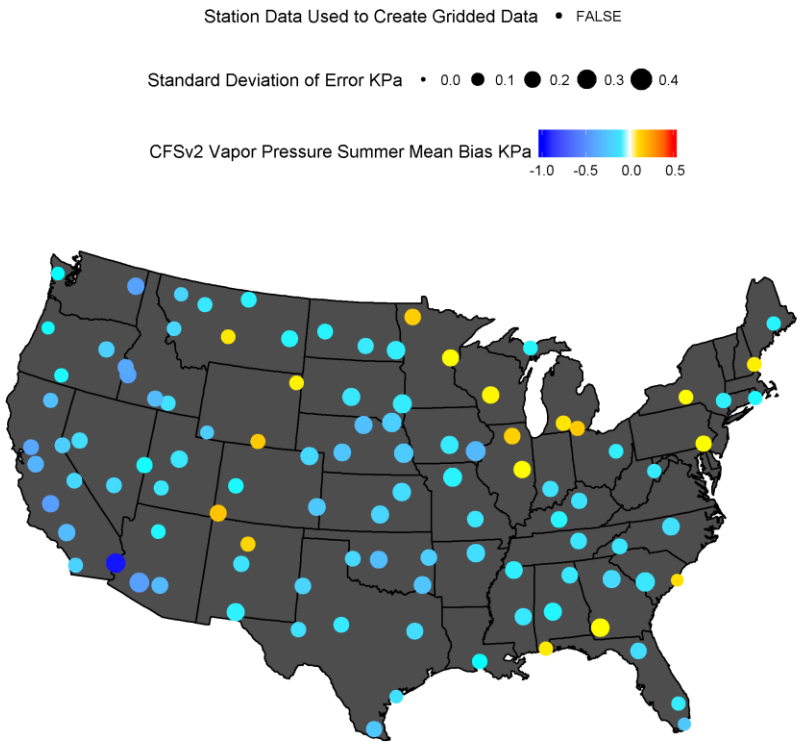

Figure 8.18 CFSv2 vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

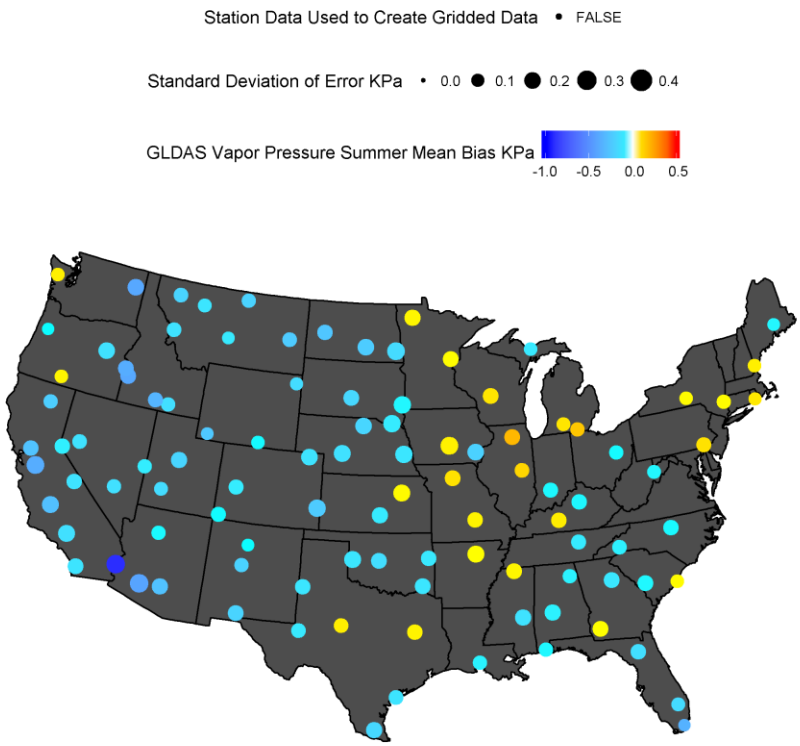

Figure 8.19 GLDAS vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

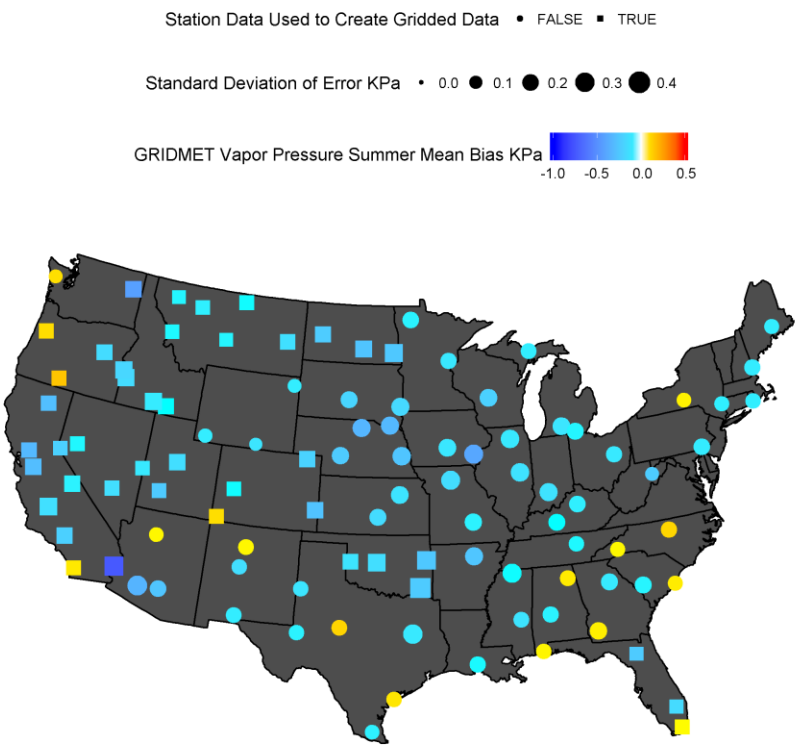

Figure 8.20 GRIDMET vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

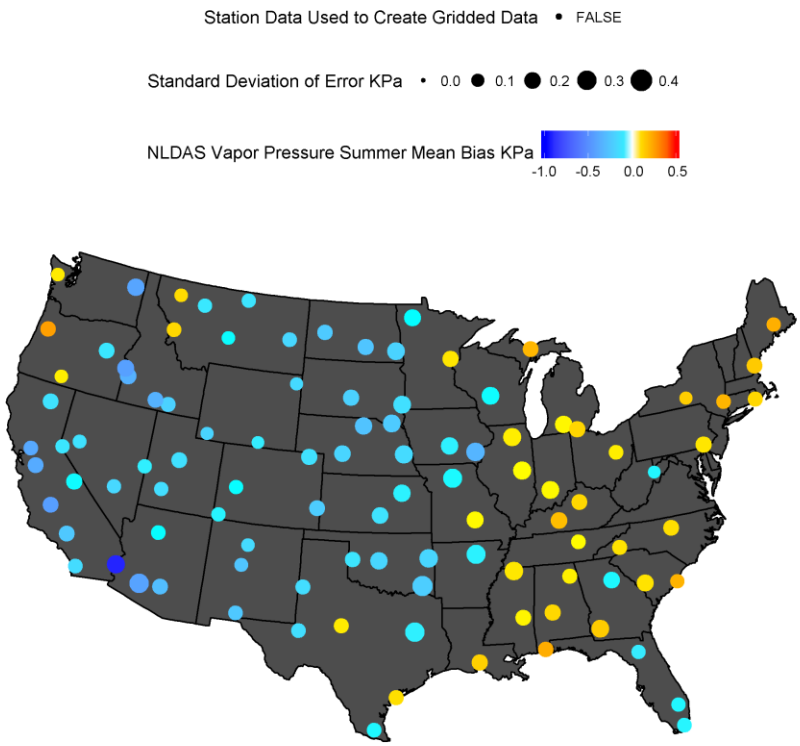

Figure 8.21 NLDAS vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

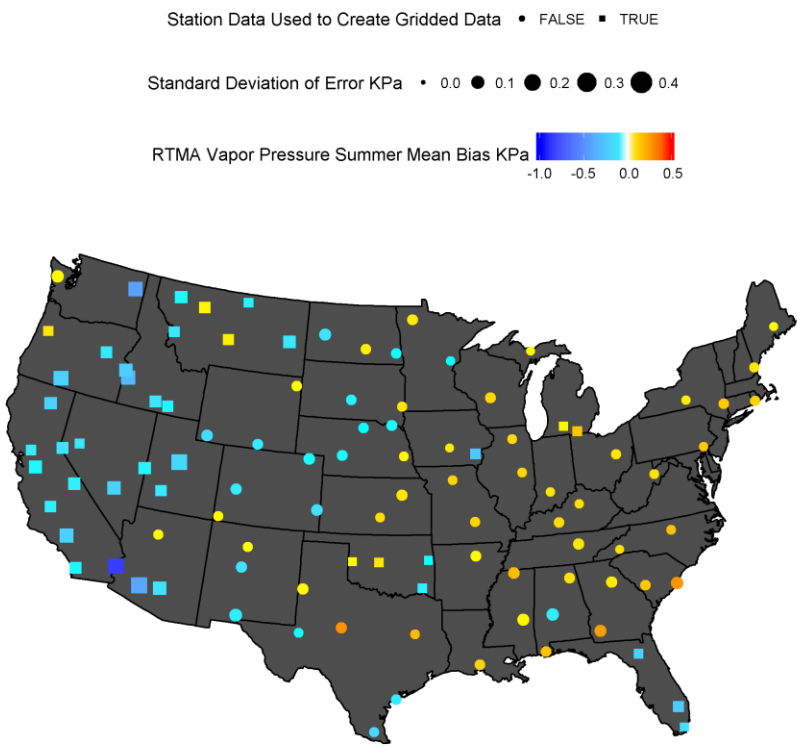

Figure 8.22 RTMA vapor pressure mean bias and standard deviation of error computed using summertime daily data for 2015.

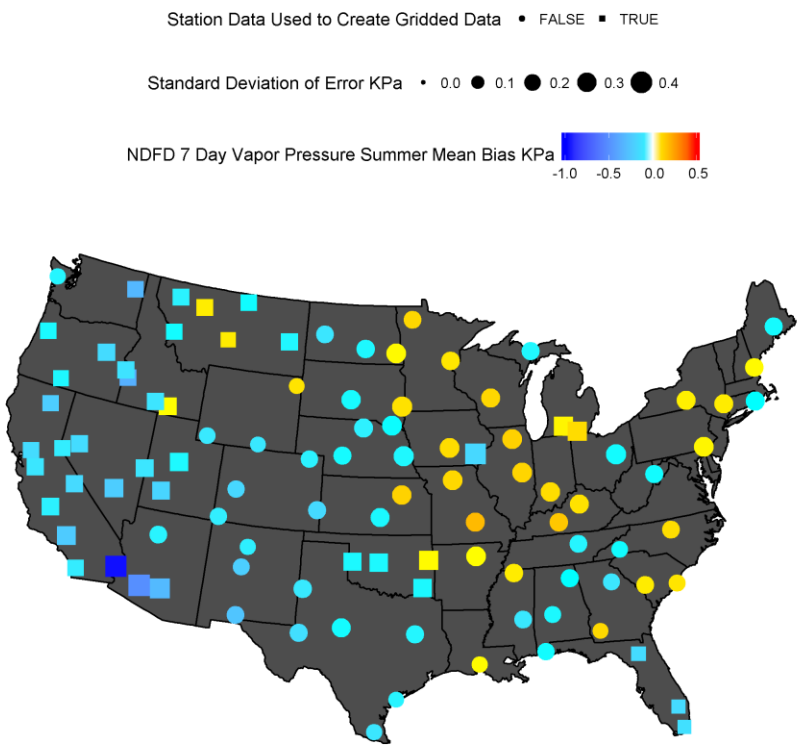

Figure 8.23 NDFD day 7 vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.

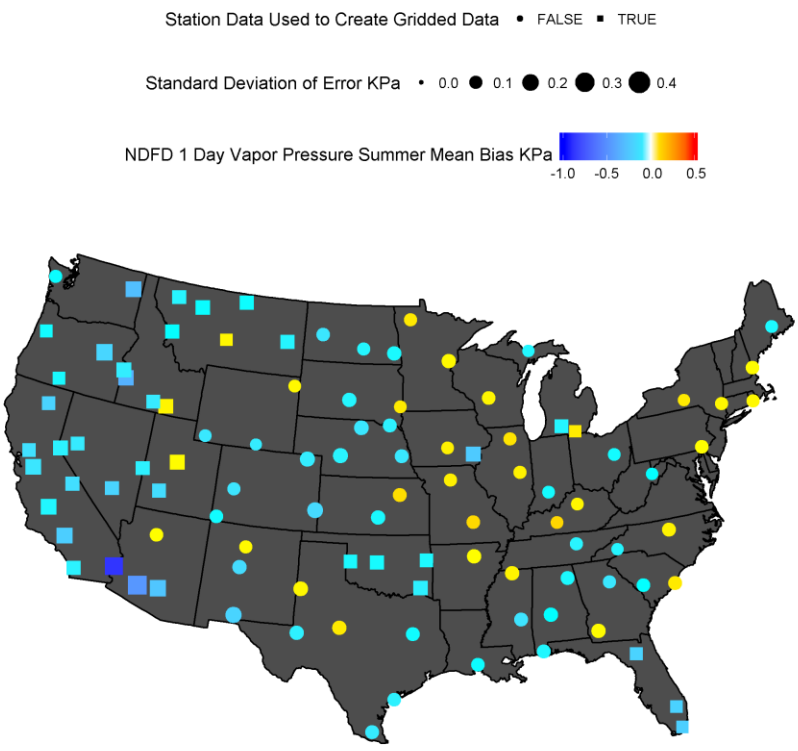

Figure 8.24 NDFD day 1 vapor pressure mean bias and standard deviation of error computed using summertime daily data pooled across all study years.