2017

An Evaluation of Gridded Temperature Products and their Effectiveness in Modeling Small Scale Ambient Temperatures

Barrie Chileen

Augustana College, Rock Island Illinois

Follow this and additional works at: http://digitalcommons.augustana.edu/geogstudent

Part of the Geography Commons

Augustana Digital Commons Citation


http://digitalcommons.augustana.edu/geogstudent/4

This Student Paper is brought to you for free and open access by the Geography at Augustana Digital Commons. It has been accepted for inclusion in Geography: Student Scholarship & Creative Works by an authorized administrator of Augustana Digital Commons. For more information, please contact digitalcommons@augustana.edu.
AN EVALUATION OF GRIDDED TEMPERATURE PRODUCTS AND THEIR EFFECTIVENESS IN MODELING SMALL SCALE AMBIENT TEMPERATURES

by

Barrie Victoria Chileen

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor’s of Arts in Geography

AUGUSTANA COLLEGE
Rock Island, IL

February 2017
# Table of Contents:

- Introduction.................................................................................................................3
- Study Area.....................................................................................................................4
- Derivation of Gridded Products......................................................................................7
- Applications of Gridded Temperature Products..........................................................8
- Current State of Gridded Product Performance..............................................................9
- Introduction to Methodology.........................................................................................11
- Data Acquisition: Control Points..................................................................................12
- Data Acquisition: Gridded Data Products....................................................................14
- Statistical Analysis.........................................................................................................14
- Results............................................................................................................................16
- Discussion.......................................................................................................................23
- References......................................................................................................................26
Introduction to Gridded Temperature Products

The use of gridded temperature products is becoming increasingly prevalent in ecological research due to their accessibility, low cost, and spatial and temporal coverage (Albright et al. 2010). While previous studies have compared gridded products against each other and weather station data, little research exists that attempts to verify the accuracy of these gridded products on finer spatial scales in field settings (Behnke et al. 2016, McEvoy et al. 2014, Holden et al. 2011). As these products are more commonly used to model climate change scenarios and understand microclimates at 2 m above the ground, it is important to note that biases in these gridded products could create a false picture of biological climate scenarios and misinform management decisions.

Gridded temperature products are derived from weather station data. Then, by using weather stations as data points, a series of algorithms are applied that factor in physical attributes (elevation, slope, precipitation) to estimate site specific temperatures. These values are then interpolated over grid cells to provide daily minimum and maximum temperatures at any given pixel on the grid. It is because of this wide scale coverage that many often seek out gridded products in place of onsite observations as onsite observations are often costly and hard to come by for a specific location. For this study, we will be looking at three widely used gridded products at varying spatial and temporal resolutions: NLDAS2 (Mitchell et al. 2004) 10 km at hourly intervals, PRISM (Daily et al. 1997) 4 km at daily intervals and Daymet (Thornton et al. 2014) 1 km at daily intervals. These products are commonly used in modeling hydrology, ecology, biology, climatology and meteorology (Albright et al. 2010, McEvoy et al. 2014) and can heavily impact management decisions that rely on these types of studies. For the purposes of
this study, we will be assessing how well these products match independently obtained ground-truthed temperature values.

The goal of this project is to evaluate and compare the performance of gridded products in modeling ambient temperatures when compared to field-determined temperature data. To do this, LogTag Trix-16 sensors were deployed in two study sites: Great Basin National Park and Kofa Wildlife Refuge in 2 meter ambient conditions. In comparing these onsite temperature sensors to the temperature readings provided by the gridded products, I hope to model potential discrepancies between data products, quantify the biases that could be associated with the spatial and temporal resolution of gridded data products, and assess how gridded product choice can impact threshold-based biometeorological indices.

**Study Area**

The study areas for this project consisted of two separate sampling sites: one located in Kofa National Wildlife Refuge in Southwestern Arizona (Figure 1) and the other site was located on the Snake Range in Great Basin National Park in Eastern Nevada (Figure 2). These two sites were chosen for two different related studies (Mutiiibwa et al. 2015, Albright et al. 2017), however they both share attributes that made them favorable for comparing the accuracy of gridded ambient temperature products. Kofa National Wildlife Refuge is classified as a Sonoran Desert and features an elevation range of 401 meters to 1,493 meters while the Snake Range site was classified as Great Basin desert with an elevation range of 1,700 meters to 3,500 meters. Both sites had a series of sensors deployed in radiation housings in ambient conditions at 2 meters above the ground in direct sunlight. These two environments featured a variety of microclimates which had made them favorable for this study: Kofa had a relatively uniform elevation profile whereas the Snake Range features sensors at a series of elevational gradients. It
is through these traits that it is possible to compare the effects of elevation range/microclimates on the accuracy of gridded temperature products.
Figure 2 Snake Range Site with Sampling Points
Derivation of Gridded Products

One of the key aspects in understanding how gridded data products perform is to understand how they are derived. Daymet, PRISM and NLDAS2 all use different parameters as input to interpolate the data observations (Table 1). Some are physical landscape attributes, for example, PRISM and Daymet both use elevation as a parameter in interpolation (Daily et al. 2008 & Thorton et al. 1997). Both PRISM and Daymet also use daily observations from a network of weather station data to provide daily temperature measurements which are then interpolated over the landscape (Daily et al. 2008 & Thorton et al. 1997). PRISM and Daymet provide daily minimum temperatures and daily maximum temperatures (Daily et al. 2008 & Thorton et al. 1997). NLDAS2, in comparison, uses the NCEP North American Regional Reanalysis (NARR) dataset and provides temperature data in hourly increments. NLDAS2 does not incorporate direct input data from elevation or local weather stations (Behnke et al. 2016).

The gridded products also have varying spatial resolutions of pixels (Figures 1 & 2): Daymet at 1x1 km pixels, PRISM at 5x4 km pixels, and NLDAS2 at 14x10 km pixels (Behnke et al. 2016). It is also important to note that these data products are to be treated as estimations of site temperatures and all observations are interpolated from weather station data. These discrepancies in data derivation, spatial resolution, and temporal resolution all have a large influence on the performance of these gridded products.

Table 1: Gridded products and derivation methods

<table>
<thead>
<tr>
<th>Gridded Temperature Product</th>
<th>Spatial Resolution</th>
<th>Daily Tmax Value</th>
<th>Hourly Temperature Value</th>
<th>Interpolation Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daymet by Oak Ridge National lab</td>
<td>1 x 1 km</td>
<td>Native to dataset</td>
<td>Cosine with Sun/Chillr</td>
<td>Elevation + Weather station data</td>
</tr>
<tr>
<td>NLDAS2 by NASA</td>
<td>10 x 14km</td>
<td>Native to dataset</td>
<td>Native to dataset</td>
<td>NCEP North American Regional Reanalysis (NOAA)</td>
</tr>
<tr>
<td>PRISM by Oregon State</td>
<td>4 x 5 km</td>
<td>Native to dataset</td>
<td>Cosine with Sun/Chillr</td>
<td>Elevation + Weather station data</td>
</tr>
</tbody>
</table>
Applications of Gridded Temperature Products

Despite these discrepancies in gridded product performance, they are still commonly used in biological and meteorological research. There have been a wide variety of applications of gridded temperature products to tackle research topics from exploring the historic relationship between drought and wildfire occurrence (Riley et al. 2013) to looking at temperature impacts to invasive shrub resistance and hybridization (Williams et al. 2014). Other studies have looked at quantifying the impacts of climate change on species abundance and species distribution (Ackerly et al. 2010). When studies use gridded products, the gridded product data is often implemented into calculating ecological indices (Riley et al. 2013). The results of these indices are then used to guide management decisions (McEvoy et al. 2013). Gridded products are often used without comparison to microclimate conditions (Ackerly et al. 2010) and their performance at these scales is often accepted to be accurate. When these gridded products are used without ground truthing, there is a risk of creating false understandings of biological and ecological scenarios. Some studies that use gridded data products fully acknowledge biases and even add justification to the use of their chosen product (Ackerly et al. 2010), however the case is often that careful selection of data products is overlooked (Williams et al. 2014). These gridded products, even with their inherent biases and shortcomings, are still one of the main products used to provide valuable climate data for studies that lack the funding or resources to go into the field and record these types of observations. It’s important to understand how they perform. There has been some work in evaluating the performance of gridded data products; however, the current understanding of gridded product performance is limited to a macroscale. In order to look into gridded product performance at microscales, it is necessary to understand the current state of gridded product performance.
Current State of Gridded Temperature Product Performance

Gridded climate datasets are becoming increasingly prevalent in biological research (Behnke et al. 2016, Holden et al. 2013, McEvoy et al. 2014) and, therefore, there is an increase in the need to understand how well these products perform. Previous studies have looked at gridded temperature product performance (Behnke et al. 2016, McEvoy et al. 2014) however; they have focused on modeling climates at large spatial scales instead of microclimates. The general understanding is that these products, though useful, have their limitations (Behnke et al. 2016). Two recent studies have assessed the performance of gridded data products at these macro-scales: The McEvoy et al. 2014 study on precipitation and temperature products along the Nevada Climate-Ecohydrological Assessment Network (NevCAN), and the Behnke et al. 2016 study on downscaled weather data performance.

The McEvoy et al. study provides a benchmark for comparison of gridded product performance as it was conducted at the Snake Range sample site. The McEvoy et al. study differs from this research project because it looks at the NevCAN monitoring network (Mensing et al. 2013) as the sampling point of comparison whereas this research project uses sensors independent of an established climate monitoring network. The main objectives of the McEvoy et al. study was to understand the differences between gridded data products and assess the potential challenges in using these gridded products to model complex terrains, which in this study referred to large topographic ranges. When looking at the Snake Range transect, it’s important to note that both Daymet and Prism share the same high elevation station as a control point for the algorithms used to derive their data (McEvoy et al. 2013). In terms of product performance within complex terrains, daily minimum temperatures are the most challenging to capture (Holden et al. 2011). Daymet had lower overall higher R² for both maximum
temperatures and minimum temperatures at the Snake Range site, which could potentially provide insight in interpreting the results of this study. Smaller spatial resolution also leads to less bias in modeling both maximum and minimum temperatures when compared to NevCAN (McEvoy et al. 2013). The McEvoy et al study provides insight on how these gridded product datasets perform in complex topographic environments and discusses the need to further “ground-truth” these products in the future at smaller scales.

The Behnke et al. study takes a different approach and focuses on assessing how well gridded temperature products perform when compared to localized weather stations. The Behnke et al. study downscaled gridded climate data and provides valuable insight as to how these products performed when compared against each other. The Behnke et al. study also includes the three data products that will be used in this research projects: Daymet, NLDAS2 and PRISM, which provides a great point of comparison in assessing gridded product performance. The Behnke et al. study lists two objectives that relate to this research project which are to assess how accurately gridded products capture temperature and to understand how spatial resolution impacts product accuracy. Daymet was found to have the smallest mean bias while NLDAS2 overestimated minima and underestimated maxima (Behnke et al. 2016). Areas with large topographic ranges were found to have the greatest discrepancies in data observations (Behnke et al. 2016). NLDAS2 was found to have the greatest discrepancies in data accuracy and was found to have the worst overall match. The Behnke et al. study establishes that gridded climate data does have similar limitations as the McEvoy et al. study since they both found that gridded climate product performance suffers in areas with large topographic ranges and that Daymet tends to outperform all other data products.
Though these two studies provide valuable insight to the current performance of gridded temperature products, they both overlook gridded product performance at microclimates, or spatial areas less than 1 km. Ecological research is often conducted at microclimates and studies involving temperature thresholds are turning towards gridded temperature products to fill in the gaps of on-site observations (Ackerly et al. 2010). In order to ground-truth these spatial gridded data products, it’s necessary to assess their performance against independently-collected data at a series of microclimate observation sites. In doing this, gridded temperature products will be evaluated at the small spatial scales that scientists require for research which could provide valuable information as to how well these products perform at the small scales they’re often used to cover.

**Introduction to Methodology**

It is already understood that there have been other studies that have assessed the performance of temperature products (Behnke et al. 2015, Holden et al. 2013, McEvoy et al. 2013). In conducting research on gridded data products and temperature observations, there is a methodological approach in comparing series of temperature observations against an established control (Chai and Draxler, 2014). This commonly used framework can be found in the Behnke et al. study, the Holden et al. study and the McEvoy et al. referenced earlier. By using similar metrics of comparison as the three studies referenced above, this study will meet the standard set of methodologies and approaches used to test the accuracy of temperature products. In doing this, it will be possible to create not only new knowledge within this field but also carry on a continuation of previously established methodologies to ensure clarity and relevance of data analysis.
The Behnke et al study, the Holden et al. study and the McEvoy et al. study all contain an approach that involves the same three key factors: a control group of data, a sample set of data and a series of statistical means of analysis. This study will contain each of these factors: A control group of on-site sensor observations, a sample group of three gridded product datasets and the standardized methods of statistical analysis. This study will analyze the data through linear regressions (R² tests) and root mean squared error (RMSE) which have been proven to be informative and commonly used metrics in assessing temperature data (Behnke et al. 2016, Holden et al. 2013, McEvoy et al.2013, Chai and Draxler, 2014 ). By incorporating the three key factors found in related studies, this project will then be able to successfully contribute to the current dialogue on gridded product data and contribute to the understanding of product performance at differing spatial scales.

Data Acquisition: Control Points

To successfully evaluate the performance of gridded temperature products, a network of temperature sensors were deployed in various microclimates to act as a control in data analysis. The control data used in this study originated from two separate, three-year studies focusing on quantifying temperature variations in microclimates (Mutiibwa et al. 2015; Albright et al. 2017). In these studies, LogTag Trix-16® temperature sensors (Figure 3) were deployed in radiation shields (Holden et al. 2013) and recorded hourly temperature values from May 2013 to July 2016. (Figure 4) For the Kofa National Wildlife Refuge Site, sensors were deployed by researchers at the University of Nevada Reno in a variety of vegetated and non-vegetated riparian sites. (Figure 4) From this dataset, a subset of sensors deployed at 2 meters above the ground in ambient conditions was used. For the Snake Range Site, sensors were deployed at varying elevations and aspects and sensors that were deployed at 2 meters above the ground in
ambient conditions were used in this study. The dataset from these samples manifests in the form of hourly temperature readings on dates ranging from January of 2014 to December of 2015 from a series of microclimates. Each microclimate is given a specific identifier that ties it to a specific site and array which, for the purposes of this study, is 2 meters at ambient conditions.

The timespan of this study focuses primarily on points taken from May 1st to September 30th of 2014 and 2015. This timespan was chosen as it best captured summer conditions and pre-snowfall data points. This becomes particularly important when working with the Snake Range dataset as the sampling sites were often overcome with snow in the winter months which impedes the sensors ability to take temperatures at ambient conditions. The spring months were also avoided due to the time required for snowmelt. The total amount of observations compared is 58,752 points for Kofa and 58,752 points for Snake Range. These observations were then compared to a series of wildly used gridded temperature products to determine their accuracy.

Figure 4 Temperature sensors (far right) deployed in radiation shields at Kofa site (left) and high elevation Snake Range site (right)
**Data Acquisition: Gridded Data Products**

The three gridded data products used in this study are NLDAS2 (Mitchell et al. 2004) 10 km at hourly intervals, PRISM (Daily et al. 1997) 4 km at daily intervals and Daymet (Thorton et al. 2014) 1 km at daily intervals (Table 1). These products were acquired from online databases (see citations above) and matched to the study sites by geographic coordinates. Pixel values were best matched to the geographic coordinates of the sensors; however, due to the spatial resolution of the various products, there was some repetition of weather station observations. For example, when using the online database for NLDAS2, there were only two pixels used for Great Basin while PRISM used four pixels and Daymet used seven pixels (See Figures 1&2 for scale). Gridded product observations were taken for the periods of May 1st to September 30th of 2014 and 2015. The total amount of observations is 58,752 points for Kofa and 58,752 points for Snake Range each for NLDAS2, PRISM and Daymet.

**Statistical Analysis**

To quantify the accuracy of gridded temperature products, this study used statistical methods to compare the control group (Logtag datapoints) against the gridded temperature data products. As discussed earlier, this project utilized standardized statistical means of analysis which are commonly used in related studies on temperature products. This study evaluated product accuracy at capturing daily minima and maxima as well as hourly values.

As mentioned previously, the gridded data products manifested in different temporal scales: NLDAS2 provided hourly temperature values while PRISM and Daymet provided daily maximum and minimum temperature values. To compare these products at an hourly scale, it
was necessary to convert daily minima and maxima into hourly values. To do this, a variety of methods were compared that used cosine functions to fill in missing hourly values based on daily minima and maxima. The interpolation methods compared were the R-Chillr function (Chillr), which is a data package in R that applies cosine curves based on longitude (Luedeling 2016), and three different methods cited in Schaub’s paper on establishing a method for estimating hourly temperatures (Schaub 1991). The three methods from the Schaub were then compared to sensor collected hourly values to determine which performed best based on linear fit, R^2 values, and RMSE. The method that performed best was a cosine fit that incorporated variable sunrise (Cossun). By applying these two interpolation methods to the PRISM and Daymet datasets, it is possible to derive hourly temperatures from daily minima and maxima to compare how products perform at hourly temporal resolutions. In order to compare daily minima and maxima, the process was much easier as the NLDAS2 dataset already had both maxima and minima native to the dataset.

Once the gridded data products were all converted to their respective temporal resolutions, a series of statistical analyses were conducted in R statistical software (R Core Team 2016) using the stats package (R Core Team 2016). The first step to analyze the data was to create a series of R^2 plots for both hourly temperatures and daily maximum temperatures. Each plot compares sensor data against gridded temperature product data and includes a line of best fit and a 1:1 line of fit. The 1:1 line of fit provides a reference point for a perfect correlation while the line of best fit represents the correlation of the actual data which then allows for the data to be easily compared. If the 1:1 line of fit is close to the line of best fit, then that would indicate that gridded product observations closely fit sensor microsite observations. The plots also include an R^2 value which represents how closely the data follows a regression line. A higher R^2
value, for the purposes of this study, would then indicate a strong fit between gridded product data and sensor data. R² values and graphical representations of both daily temperature maximums and hourly temperatures provides insight as to how well gridded products represent ambient temperatures 2 meters above the ground. These methods of analysis also would provide insight as to which hourly interpolation methods create the best fit for future research.

In related studies, there is an additional method of analysis that is consistent throughout literature pertaining to quantifying the accuracy of temperature data. Root mean squared error (RMSE) is a common metric to evaluate climate and meteorological research studies (Chai & Draxler 2014). Other papers have employed mean absolute error in addition to RMSE to test models, however, they are similar and show the similar results in data analysis (Chai & Draxler 2014). They differ in the face that RMSE gives heavier weight to errors which allows for it to better represent model performance differences (Chai & Draxler 2014). Since this study is comparing how well a sample of climate modeling data (gridded products) compares to our control data (temperature sensors) RMSE will be used instead of mean absolute error because it provides a picture of the amount of error in the predicting model (gridded products). RMSE is measured on a different scale from R² and the higher the value is, the more error exists in the model (Chai & Draxler 2014). Low RMSE values indicate that the model data fits closely to the control data which would then indicate how well gridded products are able to capture site temperatures.

Results

Data Management

Data were formatted into a series of excel spreadsheets, one workbook was established per site (Figure 5). Each workbook then contained both hourly and daily maximum and
minimum observations for each data derivation method. Tabs were created for Sensor 2014, Sensor 2015, Daymet, PRISM, NLDAS, Daymet Cossun, PRISM Cossun, Daymet Chillr, and Prism Chillr. The data in the spreadsheet covered the study period from May to September of both 2014 and 2015 with each of the eight sensors per site represented. Each of these spreadsheets were also created to be compatible with R programming software to assist in statistical analysis.

![Figure 5 Layout of raw data](image)

Figure 5 Layout of raw data
Linear regressions

Bi-variate plots with linear regression lines were created using R statistical software (Figure 6). These linear regressions were used to compare each sensor against each gridded product to see how well the model (gridded product data) fits to the control set of observations (the sensors). A line of best fit for the data was created (in green) and a line of a perfect 1:1 fit was also included as a point of reference for a perfect model. This methodology was also used to create linear regression plots for maximum temperature, however, it’s more challenging to see trends due to the considerably smaller number of points of comparison. There was a total of 48 plots per year for hourly observations, with the CosSun fit used for 24 of the plots and the Chillr fit for hourly values used for the remaining 24 plots. For daily maximum temperature, there was a total of 24 plots per year since only NLDAS, Daymet and PRISM were being compared. The R² values were also represented on these charts as a quick reference point of goodness of fit.
These hourly $R^2$ values were then summarized into boxplots for both daily maximum temperature and hourly temperature values (Figures 7&8).

**Figure 7** Hourly $R^2$ boxplots for both daily maximum temperature and hourly temperature at Kofa site

**Figure 8** Hourly $R^2$ boxplots for both daily maximum temperature and hourly temperature at Snake Range site
For daily maximum temperature at the Kofa site, $R^2$ values ranged from .56-.98. For Kofa, Daymet and NLDAS2 outperformed PRISM, with Daymet having the largest $R^2$ value while PRISM had the lowest. The average $R^2$ value for Daymet was .96, while NLDAS2 had an average of .92 and PRISM had an average of .69. For hourly temperature values, a different trend is seen. $R^2$ values ranged from .74 to .92, with NLDAS2 having the lowest value and Daymet with the Cossun interpolation method having the highest. The highest average $R^2$ value was Daymet with the Cossun interpolation, while PRISM with a Chillr interpolation method had the lowest average. At the Kofa site, Daymet appeared to have the best performance based on $R^2$ values and had a better overall fit with the control sensor collected points. PRISM had the lowest $R^2$ average for both hourly and daily maximum observations.

The data collected at Snake Range showed similar trends. Average overall $R^2$ values were highest for Daymet at .92, followed by NLDAS2 at .84 and PRISM at .71. The highest $R^2$ value was reported from Daymet at .97, while PRISM had the lowest $R^2$ at .61. For hourly temperatures, there was a much larger range of $R^2$ values, with values ranging from .20 to .90. NLDAS2 had the lowest hourly $R^2$ value while Daymet with the Chillr fit had the highest. The highest average $R^2$ was observed from Daymet with a Cossun fit, while NLDAS2 had the lowest average fit with an $R^2$ value of .37. For the Snake Range site, Daymet outperformed all other products based on average $R^2$. NLDAS2 had the worst performance when it came to hourly values, while PRISM had the worst overall fit when looking at daily maximum temperatures. RMSE followed similar trends as $R^2$, Daymet has the best overall performance for both hourly and daily maximum temperature values. PRISM had the worst performance for daily maximum temperatures while NLDAS2 had the worst performance for hourly temperature values. The same general trends apply as lower RMSE indicates less error in fit.
Another method to compare temperature sensors was to plot each gridded product on the same graph with the sensor data to act as a point of reference (Figures 9&10). A series of line charts with each gridded product in a different color to compare allowed for a visual and accessible approach to interpreting the data results. Boxplots are also shown next to each line chart to show the general distribution of data points for daily maximum temperature from June to September 2015. These plots provide a great visual representation of how these gridded data products vary from the on-site temperature readings provided by the sensors.

The sensor chosen for Kofa showed that gridded product observations generally underestimated temperatures. The boxplot distribution shows that Daymet had the closest estimate for daily maximum temperature, followed by PRISM and then by NLDAS2. The line plot also shows that gridded products tend to overestimate lower temperature readings and underestimate higher temperature readings. The Snake Range site shows a different trend as all gridded data products appear to overestimate the sensor observations. Daymet was still the
closest to matching the sensor observations, followed by PRISM and then by NLDAS2. The discrepancies between the Snake Range site and the Kofa site could potentially be caused by topography of the Snake Range site as sensor D4P8 was at one of the higher elevation points in the Snake Range sampling site.

Figure 9
Left: Daily maximum temperature observations are plotted by date for the Kofa Site
Right: General distributions of temperature observations by gridded product are shown

Figure 10
Left: Daily maximum temperature observations are plotted by date for the Snake Range Site
Right: General distributions of temperature observations by gridded product are shown
Discussion

During this study, this data suggests that there were discrepancies in gridded product choice and their performance in capturing microclimate temperatures. After looking at the $R^2$ distributions, RMSE, and $t_{max}$ plots, this data suggests that Daymet tends to outperform all other data products and that NLDAS2 has the worst performance. Though these trends align with our current understanding of gridded product performance as outlined in McEvoy et al. 2014 and Behnke et al. 2015, the scope of this study is quite limited and the results must be assessed with caution.

Some limitations of this study were that the control observations were acquired from two previously existing studies, which while useful, creates a challenge as these sensors were not deployed with the sole purpose of testing these gridded products. Another limitation is that though there were many observations, each site only ended up using eight different sensors for a total of sixteen sensors overall. Though the scope of this study was much smaller than desired, the data from this research is still valuable as it explores methods to evaluate gridded product performance at small scales.

The lasting implications of this project are seen through further applications of the data. This project is the product of one of the studies that had originally deployed the sensors. The goal of the project was to determine how microsite temperatures compare to ambient conditions to assess how bird mortality is affected by microsite availability. A side project emerged where this data was applied to a biometricological indicies as outlined in Albright et al. 2017, which in this case is identifying dehydrating degree hours over 40 degrees Celsius, to determine heat threat to birds. To identify these dehydrating degree hours, a simple evaporative water loss
analysis (Albright et al, 2017), was run in R. Instead of just using sensor data, a series of gridded products and interpolation methods were used to see how gridded product choice could influence the result of these types of biological studies. The result of this analysis is a chart that shows the number of dehydrating degree hours by each gridded product and interpolation method (Figure 11).

![Cumulative Dehydrating Degree Hours (DDH) for Block 3 Sensor 1](chart.png)

**Figure 11 Number of Dehydrating Degree Hours by Gridded Product for Kofa National Wildlife Refuge.**

The figure shows that there is a wide discrepancy between gridded product choice and the amount of dehydrating degree hours calculated as time progresses. In this scenario, NLDAS2 had the worst performance as it caps off at about 150 dehydrating degree hours, while the sensor captured over 550. The closest gridded product to the actual sensor observation was Daymet with the Cossun interpolation method. This application shows that gridded product choice can really
alter how results are interpreted, without care and caution, false biological scenarios are created and management efforts could be misguided.

While the results of this study reveal the discrepancies between gridded data product choice, much future work is needed to further test and fine tune the accuracy of gridded temperature products. As they become increasingly used in research, the greater the demand for scaled down products. Daymet had the overall best performance which could be driven by its small spatial resolution, future studies should look at other gridded products with even smaller spatial resolutions because these could provide even more insight as to how gridded temperature products perform at small spatial scales. Larger observation networks should be used as well as it provides a larger pool of data so that trends could be more easily seen. While scientists continue to use these products without properly acknowledging their potential discrepancies in data, there is a great risk that the temperature products can throw off the data results and create inaccurate results. With further testing and improved algorithms, it’s possible that someday there can be gridded products that can be used at small scales with little deviation from actual site observations.
References


Chai, T., and R. R. Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7:1247-1250.


