Journal of Geographic Information System, **2012, 4, 530-541**
http://dx.doi.org/10.4236/jgis.2012.46058 Published Online December 2012 (http://www.SciRP.org/journal/jgis)

Customized Online Aggregation & Summarization Tool for Environmental Rasters (COASTER)

Daniel J. Weiss, Katie Gibson, Charles Sojka, Jennifer W. Sheldon, Robert L. Crabtree
Yellowstone Ecological Research Center, Bozeman, USA
Email: weiss@yellowstoneresearch.org

Received August 3, 2012; revised September 5, 2012; accepted October 25, 2012

ABSTRACT

The Customized Online Aggregation & Summarization Tool for Environmental Rasters (COASTER) system (www.COASTERdata.net) was developed by Yellowstone Ecological Research Center (YERC) (www.yellowstoneresearch.org) in response to the information needs of end-user communities interested in decision-support for natural resource management. The purpose of COASTER is to greatly simplify the process of creating predictor datasets for research exploring environmental impacts driven by climate change, land-use activities, disturbance, and invasive spread. COASTER achieves this goal by providing users with a web-based system for processing environmental (gridded, raster) datasets, using a set of standardized functions, to create output customized to meet their analytical needs. In doing so, COASTER effectively translates large and cumbersome datasets into user-specified information useful for parameterizing statistical models and for visualizing spatial and temporal patterns within environmental datasets. The COASTER system currently contains over 10 terabytes of climate data from several sources. These datasets have daily temporal resolutions, spatial resolutions ranging from 1km to 330km, and temporal extents ranging from 30 to 64 years (1948-2011). COASTER datasets are primarily limited to North America, but gridded datasets from other regions can easily be added to the system. Variables within the climatic datasets available on COASTER include metrics quantifying temperature, precipitation, shortwave radiation, vapor pressure deficit, humidity, and wind conditions. Notable features of COASTER include a conceptually simple yet flexible set of functions capable of producing a wide range of outputs, a design applicable to many types of raster datasets, and results formatted for seamless integration within most GIS and remote sensing software packages.

Keywords: Geographic Information Retrieval; Online Application; Remote Geographic Data Processing

1. Introduction

Environmental variables that characterize changing habitat conditions over time are highly valuable for assessing species vital rates and community and ecosystem health, and for supporting informed resource management decisions. Variables of particular interest for such tasks include raster datasets (*i.e.*, gridded, wall-to-wall geospatial datasets) that capture climatic conditions, vegetation phenology and/or productivity, and moisture/water information with temporal and spatial resolutions sufficient for analyzing the phenomena of interest. Utilizing such variables is challenging, however, for reasons including: 1) the high cost and level of technological expertise required to produce the underlying datasets; 2) the data management capabilities necessary to acquire and store existing datasets; and 3) the computational infrastructure, software packages, and computer programming skills necessary to extract and synthesize useful information from large datasets. The Customized Online Aggregation & Summarization Tool for Environmental Rasters (COASTER) system is designed to reduce or eliminate these challenges by providing: 1) access to highly sought-after datasets, starting with interpolated climate datasets that have daily temporal resolutions; 2) a set of tools that reduce the computation and data storage capabilities required of users by processing the data on a remote server; and 3) delivering (via the web) customized products that are readily integrated, explored, and analyzed within a GIS or remote sensing software environment. The goal of the COASTER system is to produce user-defined output to meet specific analytical needs simply and easily, thereby enabling researchers with limited experience processing large geospatial datasets to include powerful variables in their analyses without requiring additional training or support.

The current (beta) version of the COASTER system contains interpolated climate raster datasets with varied spatial resolutions (as fine as 1 km), spatial extents (regional to global), daily temporal resolutions, and temporal extents (30 to 64 years in duration). The datasets

hosted on COASTER include variables from the Terrestrial Observation and Prediction System (TOPS) [1,2] for the contiguous United States and Alaska, and TOPOMET data for the contiguous United States produced by the Numerical Terradynamics Simulation Group of the University of Montana. The TOPS and TOPOMET datasets were created using similar topographically adjusted interpolation algorithms to estimate conditions in areas between meteorological base stations. The interpolation methods underlying TOPS and TOPOMET are very similar to the one used in the DAYMET model [3] and follow an approach similar to the one utilize within the Parameter Regression of Independent Slopes Model (PRISM) [4]. COASTER also contains several products distributed by NOAA including variables from the NCEP/NCAR dataset [5], the Climate Prediction Center (CPC) Unified Gauge-Based Analysis of Daily Precipitation dataset [6], and variables from the NCEP/NARR (North American Regional Reanalysis) dataset [7]. While COASTER currently contains only climatic data, the underlying algorithms are applicable to a wide range of temporally variable raster datasets including satellite imagery, classified imagery products, and modeled metrics characterizing ecosystem conditions and processes (e.g., productivity, hydrologic conditions, biomass, etc.). For a full listing of the datasets currently hosted on COASTER, including the variables each dataset contains and dataset attributes, please see http://www.coasterdata.net/documents/COASTER_metadata.html.

The datasets available on COASTER were selected because they typically must be summarized (*i.e.*, converted from daily values to some from of temporal aggregation) for use in environmental analysis and resource management applications. The daily temporal resolutions of the available datasets are also well suited for demonstrating the power and flexibility of the COASTER system. All datasets on COASTER have been validated by the organizations that produced them, but as with all modeled datasets, there are limitations that users must consider. It is important to note, however, that inherent limitations within individual datasets are beyond the scope of this paper.

2. Background

Creating and/or distributing raster climatic datasets is a challenge being undertaken by several different groups. The NOAA Earth System Research Laboratory, Physical Sciences Division (www.esrl.noaa.gov/psd/) distributes a diverse and useful set of gridded meteorological datasets at high temporal but low spatial resolutions (*i.e.*, 25 km and higher). Notable projects that produce and/or distribute data at comparable or finer resolutions include DAYMET (daymet.ornl.gov), PRISM (www.prism.oregonstate.edu) [4], Climate Western North America (www.genetics.forestry.ubc.ca/cfcg/ClimateWNA/ClimateWNA.html) [8,9], the Climate Research Unit (CRU) (www.cru.uea.ac.uk/) [10], WorldClim (www.worldclim.org/) [11], CliMond (www.climond.org) [12], and Arctic RIMS (rims.unh.edu). These projects are differentiated by the datasets they contain, the underlying models used to create the data, and the data distribution methods they employ.

What sets COASTER apart from these valuable data sources is that COASTER goes beyond distributing modeled products by allowing users to create customized products specified to meet their analytical needs. COASTER accomplishes this through a combination of the data it contains and its data processing functionality, which offers tremendous flexibility for creating user-specified datasets while minimizing the processing capabilities and data storage capacity required of end users. Another strength of COASTER lies in its applicability to a wide array of environmental datasets, as virtually any high temporal resolution raster dataset can be added to the system with relative ease. COASTER is also unique in that it is not linked to a single data producing organization and contains numerous examples of variables (from different sources) designed to capture the same or similar environmental phenomena. By offering a means of applying identical processing functionality to seemingly redundant datasets, COASTER provides users with a greatly simplified mechanism for directly comparing datasets. In other words, users can make matching products from different datasets and examine the results, in combination with the accompanying metadata, to determine which dataset is best suited to meet their needs.

3. Methods

3.1. COASTER User Interface

The online interface to the COASTER system (**Figure 1**) can be found at www.COASTERdata.net. To keep the COASTER system as streamlined and robust (e.g., equally functional in all web-browsers) as possible we opted for a relatively simple web interface in which we use a single web page for collecting all user information necessary for creating COASTER outputs. A positive side effect of this decision is the ease with which a GIS-based tool could be created to collect the user specifications, thereby providing a more graphical COASTER interface if one is desired. We also made the design decision to only collect latitude and longitude coordinates to define the region of interest (*i.e.*, the spatial subset of the dataset's full spatial extent delivered in the output file). The rationale behind this decision was our frequent past dissatisfaction with existing data distribution systems

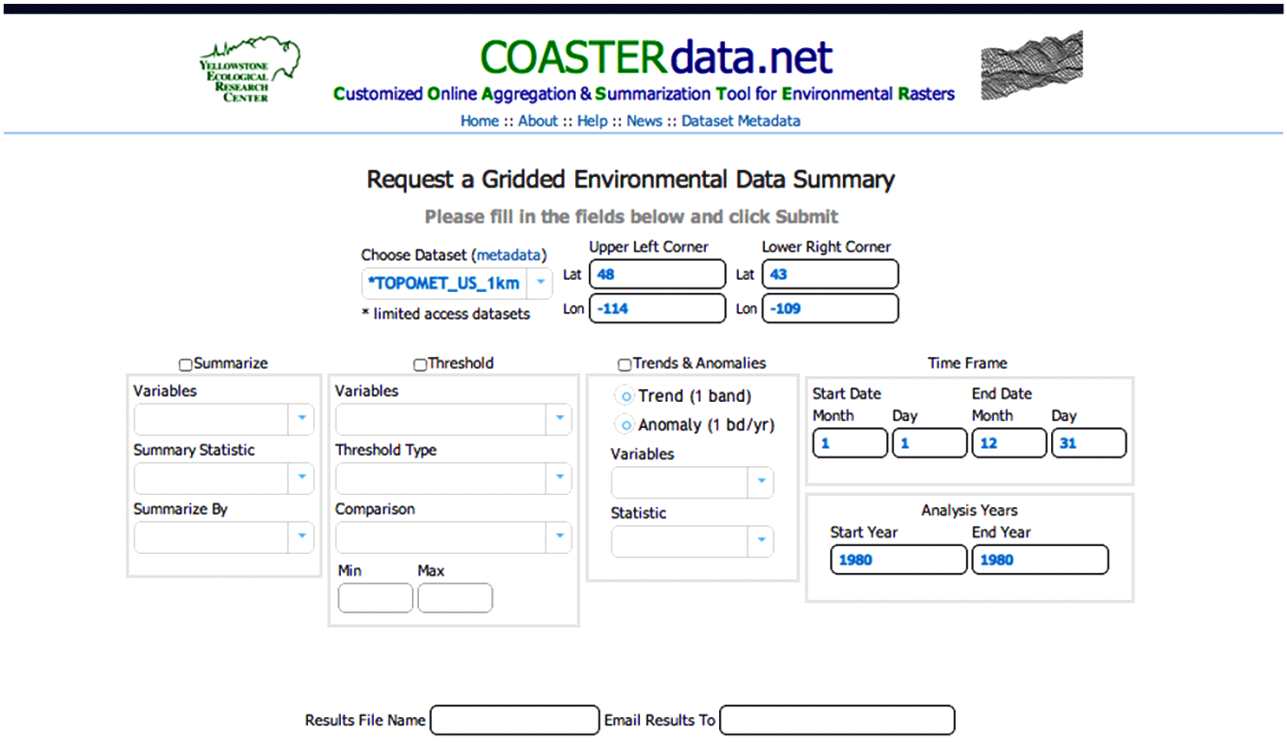

Figure 1. The online COASTER interface found at www.COASTERdata.net.

that use interactive mapping interfaces (e.g., those where users drag a bounding box over a map to specify a region of interest).

3.2. Computational Structure

The data processing elements of the COASTER system originated as a series of IDL scripts written for use with ENVI imagery processing software. To make the COASTER system stand-alone, and to avoid the licensing issues, we rewrote the COASTER data processing code in C#, taking advantage of the freely available GDAL library for processing geospatial datasets. User-specified jobs entered via the online interface are first sent to a job queue where they are stored until a processing server containing the selected dataset picks them up. All dataset and processing server details and linkages are managed using web-based administrative consoles, thereby easing the process of adding or editing datasets available within COASTER. Key challenges the architecture of the COASTER addresses include 1) creating a simple yet effective web-based user interface; 2) designing a system that places few computational demands on user's machines, since producing COASTER-like outputs locally can be prohibitive; 3) limiting file search time and I/O (Input and Output—the reading and writing of files to hard-disk) demands that slow COASTER performance; 4) creating a scalable system amenable to potentially hosting many large (e.g., up to several terabytes in size) datasets; and 5) delivering outputs that require minimal additional processing prior to being used by researchers.

A key design feature of COASTER is scalability, which is achieved through the system architecture that is designed for distributing data and processing demands across many servers, potentially located in multiple locations. The COASTER interface consists of a single web application that accepts user requests for summarized results derived from datasets available on the system. Parallel to the web interface are two database tables that 1) associate specific datasets with processing server(s) and contains details about the datasets (e.g., their data path on the host server and file naming conventions), and 2) store user submitted processing requests (*i.e.*, the job queue). The job queue serves as the linkage between the web interface and the processing servers, and user requests are ultimately executed only when a processing server claim a job from this queue. Queuing was implemented for all COASTER processing due to the intensive I/O required of many COASTER jobs, which greatly reduces data access speeds, and therefore processing efficiency, if/when a processing server attempts to run multiple jobs for the same dataset. In other words, to minimize run times, processing servers within the COASTER system process only one job at a time, and subsequent user requests are only claimed from the job queue when the processing server is free to do so. **Figure 2** shows the key components of the COASTER system and how they are linked.

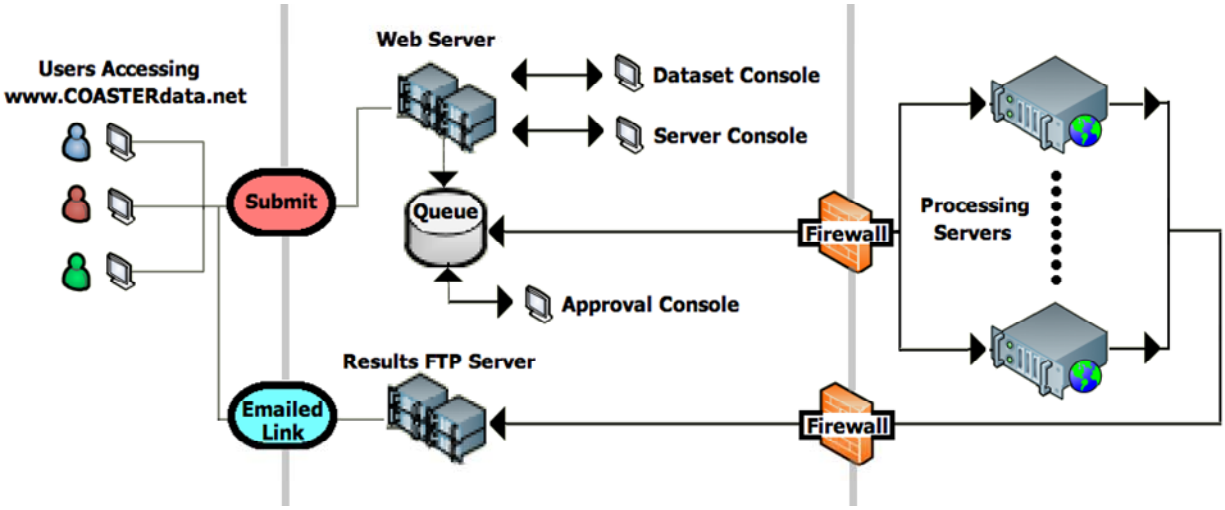

Figure 2. The COASTER system architecture diagram.

When a COASTER processing server completes a job, the resulting files are uploaded to an FTP server and the user who made the request is sent an email containing a URL link to their processed output. COASTER results consist of 1) a GeoTIFF image in the native data projection (*i.e.*, the projection associated with each dataset available within COASTER, unchanged from the projection defined when the original dataset was created); 2) information necessary for reprojecting the data as needed; and 3) a text file documenting all user-specified arguments entered on the web interface. Note that COASTER was specifically designed without a built-in reprojection utility to 1) reduce the complexity of the system; 2) avoid the pitfalls of black-box processing (*i.e.*, those in which the user has no control over, or knowledge of, the transformation procedure); and 3) maintain the integrity of the original dataset through delivery to the end-user (*i.e.*, all statistical procedures within COASTER will be unaffected by image transformations). As a result it is incumbent upon the end-users to reproject COASTER output to match their existing datasets (or vice-versa), but it is our belief that this procedure is best done with direct oversight.

The in-line processing architecture of COASTER is potentially limiting if the number of unprocessed job requests stored in the job queue becomes very large and/or if many jobs are requested from the highest spatial resolution datasets (*i.e.*, those with the largest file sizes). COASTER uses two primary mechanisms to reduce the likelihood of system slowdowns in these cases. The first mechanism relates to system scalability, as additional processing servers can be added to COASTER to meet user demand. Furthermore, COASTER supports one-to-one, one-to-many, and many-to-many relationships between datasets and processing servers, which allows 1) high-demand datasets to be mirrored across multiple servers and/or 2) very large datasets to be hosted on dedicated servers, thereby freeing other processing servers from the responsibility of running very time-consumptive jobs. The second mechanism to reduce the load on the COASTER system is by limiting access to the highest spatial resolution datasets. The majority of datasets on COASTER are freely available to the public, and jobs utilizing these datasets have a typical run-time of a few minutes. In contrast, jobs from higher spatial resolution (e.g., 1 km) datasets can take hours to process, and we therefore limit access to these datasets. When users request an output derived from a limited access dataset, they are sent an email asking for details related to the processing request, such as the purpose of the requested output and the organizational affiliation of the user. Based on the answers to these questions, the job will either be approved or denied by a YERC employee using an approval console. For users associated with organizations that have helped fund the development of COASTER, approval is essentially automatic.

The data processing functionality underlying COASTER is not highly sophisticated, as the summarization options available to users are neither computationally intensive nor mathematically complex. However, designing a system capable of efficiently opening and reading, spatially subsetting, and performing mathematical functions on all cells (e.g., potentially numbering in the millions), from thousands of raster files (e.g., the daily record for a single climatic parameter spanning 60+ years amounts to more than 20,000 files), from datasets with varied characteristics (e.g., different projections and spatial and temporal extents) was a significant computational challenge. The legacy of the COASTER system as a set of desktop tools, designed to function despite processing and memory limitations, was useful for keeping the computational infrastructure requirements of COASTER relatively low.

The result is a system that keeps at most two temporary arrays (each the size of a raster file for a single day) stored in memory as the "working tabulations" necessary for the user-selected data summarization function. When processing, each daily raster file required by the user-selected function is opened, processed, and then closed before moving onto the next file. In this way COASTER need not have all necessary data loaded into memory simultaneously, thereby enabling it to function on relatively inexpensive servers.

In addition to the online tool, the COASTER system may also be utilized as Software as a Service (SaaS), accessed using Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) based web services. This functionality allows COASTER to be accessed using an automated call to a URL whereby the user-defined parameters are sent directly to a web service. Examples of tools that could take advantage of this functionality include custom applications built within GIS environments, as well as existing online data visualization and mapping tools. A Web Services Description Language (WSDL) description is provided as part of the web service to describe how it is called and what parameters it expects. We currently do not have COASTER listed with a Universal Description Discovery and Integration (UDDI) registry, but will consider that discovery mechanism as the service is rolled out for wider use.

3.3. Adding Datasets and Processing Servers

Adding new datasets and processing servers to COASTER is accomplished through the online administrative console. To add a dataset to a processing server already running the COASTER software the procedure consists of 1) creating a new dataset record containing the necessary details within the dataset SQL database, and 2) copying the dataset onto the appropriate processing server(s). To add a new processing server to COASTER the data processing software must first be installed on that server and the server must be configured appropriately (e.g., set up as a web server running IIS, setting permissions to allow COASTER to call the necessary subroutines, and installing the queue processor Windows Service). Once the processing server is configured it is added to the COASTER system using the online administrative console. The ease of adding new datasets and processing servers to COASTER greatly enhances the scalability of the system, particularly as processing servers may be located in different physical locations.

3.4. Functions

The functions available on COASTER are divided into three categories: summarization (**Table 1**), threshold (**Table 2**), and anomaly and trend detection (**Table 3**). Each of these categories corresponds to a section (outlined by a gray box) of the user interface, within which the parameters of the function are defined. Each user request is also accompanied by 1) the temporal parameters, specified in the "time frame" section of the interface, and 2) the spatial extent of the output (*i.e.*, the spatial subsetting portion of COASTER) as defined by the upper left and lower right corner coordinates. Upon clicking the submit button all the user-specified information is stored as a new record in the job queue.

Variable definitions:

Y = the resulting value (note that all equations are applied to *individual cells*);

X = the value for a single grid cell on a single day;

$d1$ = the starting day of the intra-annual period;

dn = the ending day of the intra-annual period;

dx = the day currently being processed in an outer (loop) equation;

$y1$ = the starting year of the inter-annual period;

yn = the ending year of the inter-annual period.

3.5. Function Notes and Caveats

- All calculations done on a per year basis will produce an image with one or more bands (up to the number of years in the dataset), with each band representing the results from a single year.
- All functions currently available in COASTER are applicable only to single climatic parameters. Functions capable of processing multiple parameters are desirable (e.g., the amount of precipitation falling when the temperature is below 0 degrees C) and may be added in the future.
- Most COASTER functions are limited to a single pass through the data, meaning that functions able to summarize already summarized data are not yet available within COASTER. An example of such a dataset would be one that identifies the coldest mean temperature for the month of April from all years in the measurement period. Producing this dataset would require first calculating a per-year mean product and then processing that intermediate result using a minimum function.
- Standard deviation and Root Mean Squared Difference from Normal (RMSDN) are related functions, but they capture fundamentally different phenomena. Standard deviation raster results provide a measure of how variable a given period is relative to the mean value for that period. For example, per-year results for standard deviation compare each day to the mean value from the corresponding year to produce a measure of how variable conditions are during a single year. In contrast, RMSDN results focus on how unusual the days within a time period are relative to the daily normal values (based on the date-specific mean

Table 1. Summarization functions available within COASTER.

Summary Statistic	Summarize by	Equation	Example
Mean	All-Years	$Y = \frac{\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)}{(dn-d1)*(yn-y1)}$	The average summer (June 1st-August 31st) maximum temperature for 1980 to 2009.
Mean	Per-Year	$Y = \frac{\sum\left(X_{d1\cdots dn}\right)}{(dn-d1)}$	The average summer (June 1st-August 31st) maximum temperature for each year in the 1990s.
Standard Deviation (Std Dev)	All-Years	$Y = \sqrt{\frac{\left(\sum\left(X_{d1\cdots dn}\right)-\left(\frac{\sum X_{dx,y1\cdots yn}}{(yn-y1)}\right)\right)^2}{(dn-d1)}}$	The inter-annual variability in fall (September through November) minimum temperature, relative to the average value for all days within the measurement period.
Standard Deviation (Std Dev)	Per-Year	$Y = \sqrt{\frac{\left(\sum\left(X_{d1\cdots dn}\right)-\left(\frac{\sum X_{d1\cdots dn}}{(dn-d1)}\right)\right)^2}{(dn-d1)}}$	The intra-period variability in fall (September through November) minimum temperature in 2005, relative to the 2005 fall average minimum temperature.
Range	All-Years	$Y = \max\left\{\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)\right\} - \min\left\{\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)\right\}$	The difference between the highest and lowest maximum temperatures in all December days from 1995 to 2005.
Range	Per-Year	$Y = \max\left\{\sum\left(X_{d1\cdots dn}\right)\right\} - \min\left\{\sum\left(X_{d1\cdots dn}\right)\right\}$	The difference between the highest and lowest maximum temperatures in all December days for each year from 1995 to 2005.
Sum	All-Years	$Y = \sum\left(X_{d1\cdots dn,y1\cdots yn}\right)$	Total precipitation over all summer months from 2004 to 2006.
Sum	Per-Year	$Y = \sum\left(X_{d1\cdots dn}\right)$	Annual precipitation amounts for September from 2004, 2005, and 2006.
Min	All-Years	$Y = \min\left\{\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)\right\}$	The record minimum temperature for January 1st within the measurement period (1980 to 2009).
Min	Per-Year	$Y = \min\left\{\sum\left(X_{d1\cdots dn}\right)\right\}$	The temperature of the coldest day in the winter of 2005-2006 (November through March).
Max	All-Years	$Y = \max\left\{\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)\right\}$	The record maximum temperature for July 4th within the measurement period (1980 to 2009)
Max	Per-Year	$Y = \max\left\{\sum\left(X_{d1\cdots dn}\right)\right\}$	The total precipitation that fell during the largest precipitation event of May each year in the 1990s.
Root Mean Squared Difference from Normal (RMSDN)	All-Years	$Y = \sqrt{\frac{\left(\sum\left(X_{d1\cdots dn,y1\cdots yn}\right)-\left(\frac{\sum X_{dx,y1\cdots yn}}{(yn-y1)}\right)\right)^2}{(dn-d1)*(yn-y1)}}$	The average absolute difference between the daily maximum temperature and the average daily maximum temperature for February. RMSDN (all years) provides a measure of how unusual conditions have been throughout the measurements period (e.g., the 2000s) for an intra-annual period (e.g., October-December).
Root Mean Squared Difference from Normal (RMSDN)	Per-Year	$Y = \sqrt{\frac{\left(\sum\left(X_{d1\cdots dn}\right)-\left(\frac{\sum X_{dx,y1\cdots yn}}{(yn-y1)}\right)\right)^2}{(dn-d1)}}$	The average absolute difference between the daily minimum temperature and the average daily minimum temperature for February 1988, 1989, and 1990. RMSDN (per year) provides a measure of how unusual (relative to daily averages) conditions were for an intra-annual period within each year.

Table 2. Threshold functions available within COASTER.

Threshold Type	Comparison	Example
First Occurrence*	Greater Than	The day within the month of January, 2000 when the minimum temperature first got above −10 degrees C.
First Occurrence*	Less Than	The day of the year within the fall (August-November), 2008 when the maximum temperature first fell below 10 degrees C.
First Occurrence*	Between	The date within the entire year when the maximum temperature fell between 5 and 10 degrees C.
Count	Greater Than	The number of days within spring (March-May) when the amount daily precipitation was above 10 mm.
Count	Less Than	The number of days in spring (March-May) when the minimum temperature fell below 0 degrees C.
Count	Between	The number of days during a growing season (May through September) when the maximum temperature was most advantageous for a crop of interest (e.g., between 15 and 25 degrees C).

*The index value of the grid cell identified by the First Occurrence function represent the Julian day of the year on which the event first occurred.

Table 3. Trend and anomaly detection functions available within COASTER.

Type	Statistic	Calculation	Example
Anomaly (per-year)	All (see **Table 3**)	The per year result minus the all year result for the selected function.	The difference between the average maximum temperature for the period from April 15th to May 15th in 2005 and the 30-year (1980-2009) mean temperature for the same period.
Trend (all-years)	Mean or Sum (see **Table 3**)	The slope of the linear trend fit through the per year results for the selected function.	The linear trend in total precipitation for the month of August from 1980 through 2009. Values above zero indicate increasing precipitation over time and vice-versa.

values from all years in the dataset) for the same period. An example of a direct comparison between per-year standard deviation and RMSDN can be made by looking at minimum temperature for an 8-day period during which a bitterly cold air mass was present for a region of interest. During that time the per-year standard deviation may be very low due to fairly consistent temperatures from day to day, but the RMSDN would be very high if the daily conditions were well below the corresponding daily normal values.

- The trend analysis and anomaly detection functions are best viewed as experimental as they may be heavily influenced by the changing modeling conditions for the underlying dataset (e.g., addition of new meteorological stations and/or changing instruments that will impact interpolated climate datasets such as TOPS). COASTER trend and anomaly outputs will show interpretable trends at regional scales, but fine scale analyses (*i.e.*, analysis of a single pixel) should be done cautiously.

4. Results

COASTER presently contains over 10 terabytes of environmental data within over 700,000 daily raster files. The vast majority of this data was added in July, 2012 as version 2 of the system was unveiled. Despite not widely publicizing COASTER and the limited amount of data hosted on the system for the majority of its existence, since first becoming operational in March, 2011 www.COASTERdata.net has been visited by over 500 unique users from 40 countries, many of whom returned to the site on numerous occasions. While COASTER usage has been modest thus far, we expect (and have prepared for) a significant increase in the number of site visits due to the recent addition of so many new datasets as well as an increased effort to publicize the system.

The processing and delivery time for COASTER outputs for all but the highest spatial resolution products (*i.e.*, the 1 km, limited access datasets), even when applying the most computationally intensive functions to the longest possible time periods, is typically less than one hour. Smaller jobs, in contrast, are often completed and delivered in a matter of minutes. Delivery times increase considerably when many user requests are present in the job queue, but thus far COASTER has provided at worst next-day delivery of all requested outputs, including those derived from limited access datasets.

The greatest strength of the COASTER system lies in its ability to effectively convert massive and cumbersome amounts of *data* into *information* useful for research, education, and informed decision-making. The sheer volume of data and number of files that may require processing when dealing with high temporal resolution datasets epitomizes this capability. While the datasets available on COASTER can be daunting in their scope, the summarized products produced from them need not be, and, in fact, users are likely already quite familiar with products that can be produced using COASTER. For example, a COASTER output quantifying the mean high temperature for a single day is conceptually similar to the "normal" high temperature for that day as reported on the local news. Another strength of COASTER lies in

its ability to produce spatially specific results, with unique values calculated for each grid cell within a userspecified study area, thereby allowing spatial patterns in the newly made variables to be explored visually and statistically. **Figures 3-6** illustrate COASTER outputs displayed within a GIS (in this case overlain by a US state boundaries layer). Note that the example maps shown represents only a fraction of COASTER's capabilities, as many more functions exist than are shown.

A descriptive example of an application of the

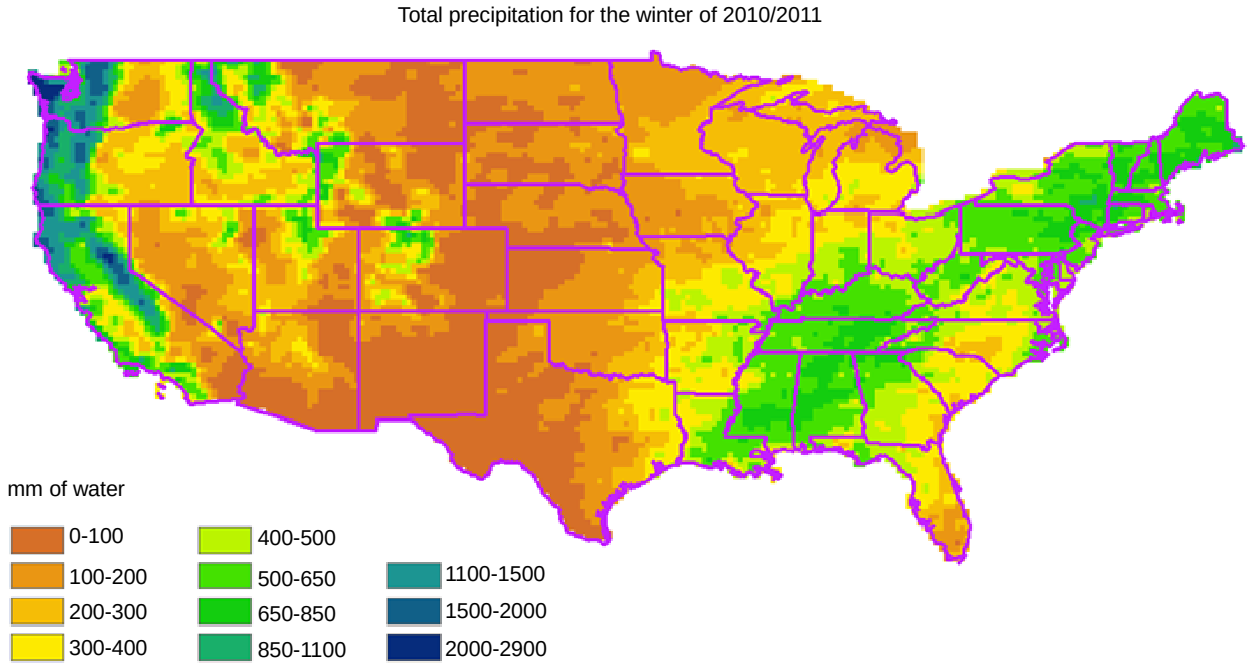

Figure 3. Example output produced using a COASTER summary function. In this map each cell value contains the sum of daily precipitation for all days between October 1st, 2010 and March 31st, 2011.

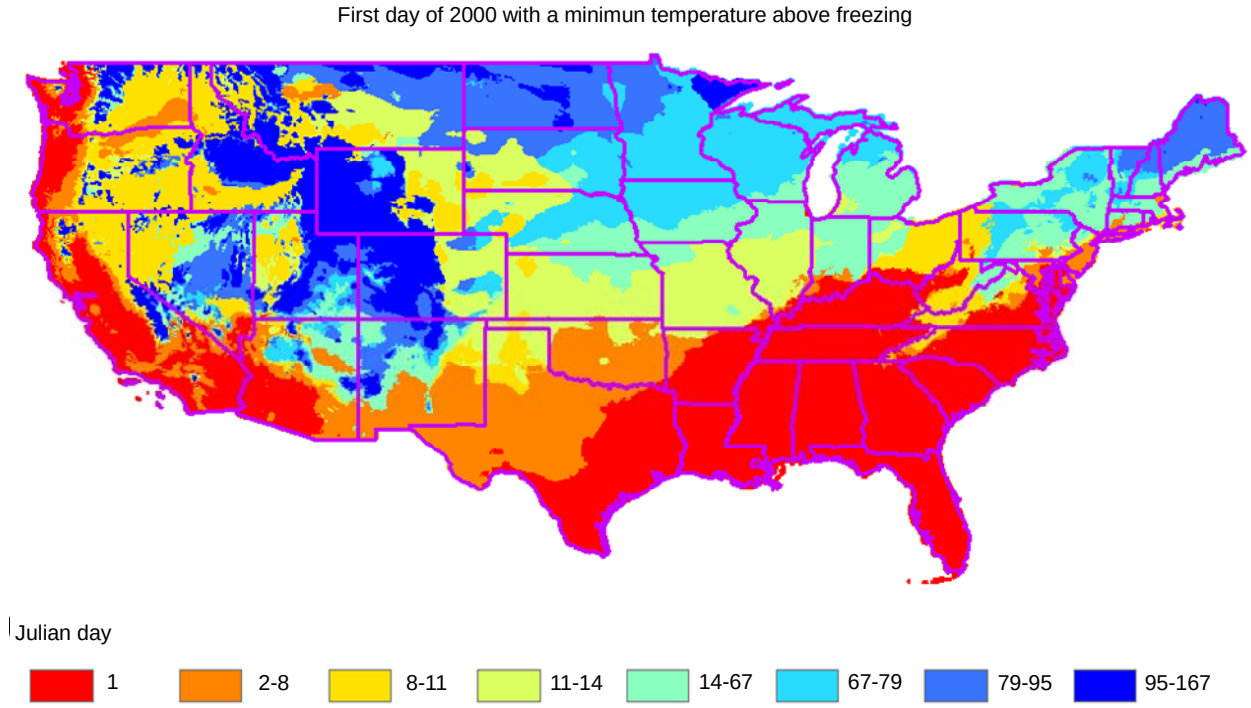

Figure 4. Example output produced using a COASTER threshold function. In this map each cell value contains the Julian date of the first day in the year 2000 when the minimum daily temperature values was above 0 degrees Celsius.

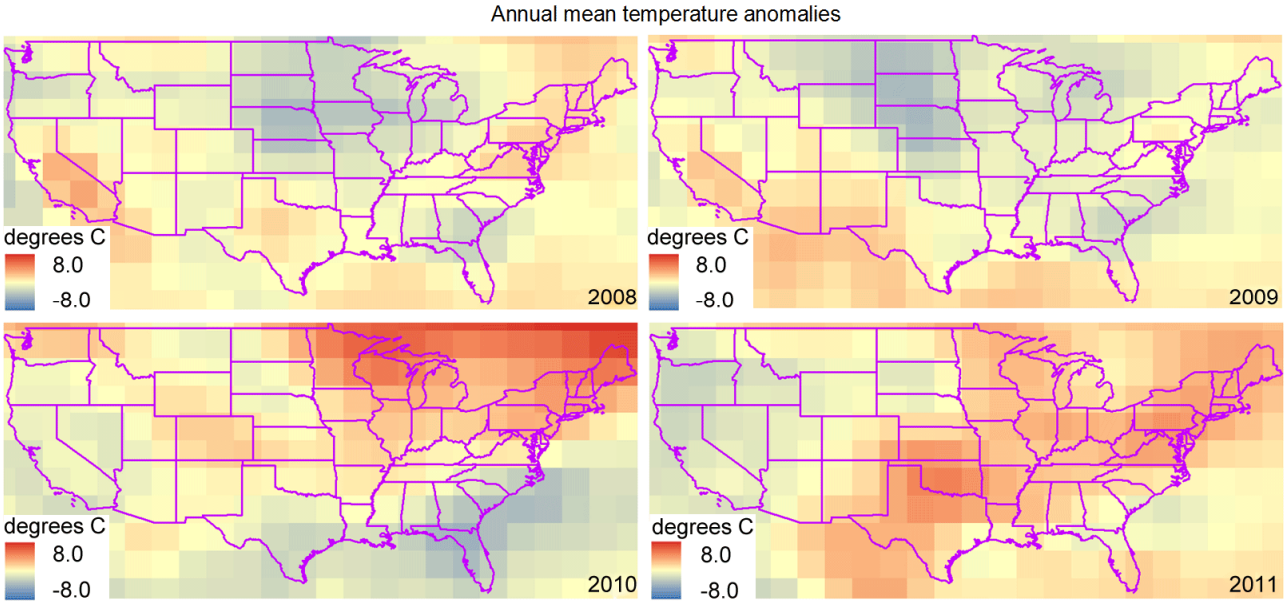

Figure 5. Example output produced using the COASTER anomaly function. Each map shows the difference between the annual mean temperature (per cell) for a specific year and the average annual mean temperature derived from all years in the dataset (1948-2011).

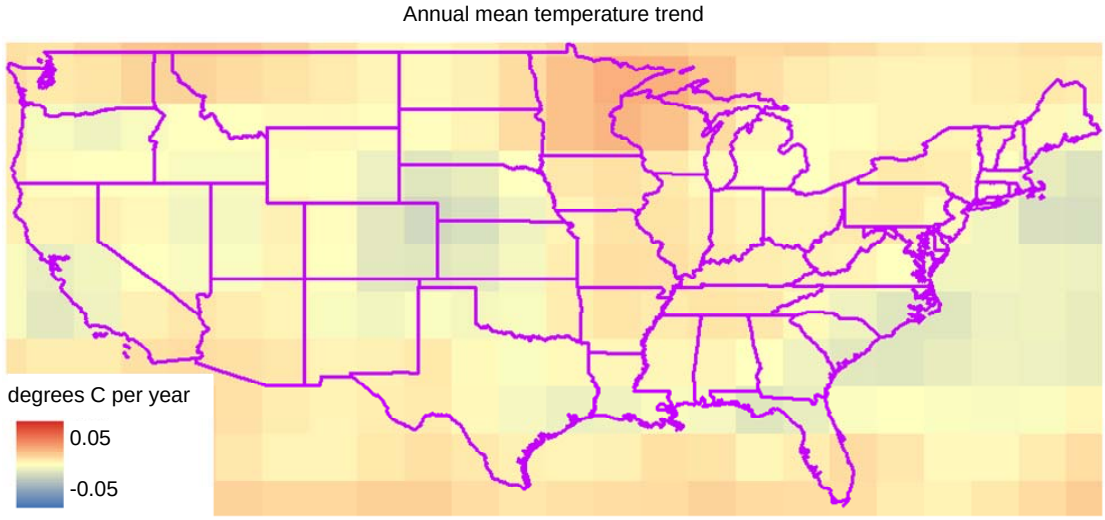

Figure 6. Example output produced using the COASTER trend function. In this map each cell shows the slope of the 64-year mean temperature trend (1948-2011) within the NCEP/NCAR dataset.

COASTER system is a hypothetical analysis of waterfowl population dynamics. Such an analysis may call for variables that quantify the climatic conditions prior to the estimated arrival time of the focal species in its summer nesting grounds, as these variables pertain to the type, abundance, and spatial pattern of available food resources on the landscape. Using COASTER it would be very simple to derive mean temperature (min and max), temperature (min and max) anomalies, total precipitation, and precipitation anomalies for multiple intra-annual temporal windows. Such datasets would enable researchers to rapidly test many hypotheses associating waterfowl observations (e.g., habitat use or nesting success.) with climatic conditions using a wide array of statistical approaches.

5. Discussion

The novelty of COASTER lies in its ability to combine features within a single online system that typically require multiple software packages and significant time and expertise to produce. Specifically, COASTER combines the following functionality to greatly reduce or eliminate many demands placed on users in need of customized raster outputs:

1) Data Distribution—relatively small output files are downloadable via a simple FTP link;

2) Remote Storage—eliminates the need for users to acquire the raw data, thereby reducing user data storage and bandwidth requirements greatly;

3) Processing Capabilities—COASTER's processing functionality combined with the ability to define regions and temporal windows of interest allow users to create a virtually unlimited number of output products without the need to develop custom algorithms and write custom scripts. Furthermore, an important difference between COASTER and other environmental raster data distribution sources is that COASTER does not limit users to pre-defined temporal summarization windows (e.g., monthly or annual summaries) that may not fit their analytical needs.

1) Remote Processing—all processing occurs remotely on the COASTER system, freeing users from needing the computational capability to generate equivalent outputs;

2) Data Formatting—all COASTER outputs are delivered as GEOtiffs that are GIS ready (*i.e.*, they are not delivered in an esoteric format that requires multiple steps prior to visualization).

To keep computational and maintenance costs low COASTER was designed to run on desktop machines and, as a result, the current COASTER system runs effectively on just a few servers. If demands on COASTER remain low, the current system may be the logical home of COASTER going forward. However, should usage of the COASTER system increase dramatically, the delivery time for completed outputs will likely increase, and a new strategy may be needed. Since COASTER provides data free of charge we feel data delivery times of several hours are not unreasonable, particularly given the time and effort it would take to create comparable results without COASTER. However, if data delivery times start routinely exceeding 24 hours we will pursue one or more the following options (pending funding availability): 1) purchase more servers to mirror the available datasets to better distribute COASTER processing demands across more nodes; 2) move to a multi-threaded processing architecture whereby we take advantage of multiple nodes to execute single jobs more quickly; 3) build collaborative relationships in which we provide the COASTER software, including access to our online interface, to other institutions (e.g., universities or governmental agencies) to host specific datasets on their servers; or 4) migrate COASTER to a cloud computing environment. All of these options are attractive because, from a user's perspective, the experience of using COASTER will be unchanged as the processing servers, wherever they may be, will simply be claiming records from the job queue. In terms of time and effort, shifting to a cloud computing architecture would be the most costly of these options. However, some of the challenges to such a transition will be eased if we host COASTER at a commercial cloud computing facility, such as Google, Microsoft Azure, or Amazon Web Services. Currently the best fit for the necessary cloud services would be with Amazon, as Google doesn't offer the type of hosting needed and Microsoft's Azure requires payment for an entire server instance and then offers up to only 2 terabytes of storage, while COASTER's requirements currently exceed 10 terabytes. Amazon Web Services would allow COASTER to rent the requisite data storage through their Simple Storage Service (S3) and to pay for processing time on an as-needed basis through their Elastic Compute Service. These and other cloud-based solutions will continued to be explored and later versions of COASTER may migrate to these platforms if/when user demand exceeds our existing processing capacity.

Data quality and continuity are concerns for the COASTER project as they directly impact the utility of COASTER for scientific research and decision-making. While we strive to provide high quality datasets, and these data are often the best or only available datasets of their kind, no dataset hosted on COASTER is without error and it is very important that users understand and acknowledge these errors in their research. To get a sense of data quality, COASTER users are encouraged to explore all metadata and read the scientific literature associated with the input datasets they select prior to creating and using the summarized outputs. Data continuity is another challenging issue for the COASTER project as we do not produce the datasets hosted on the system, and are therefore only able to update datasets if/when such updates are made by the primary data providers. Fortunately several of the datasets available on COASTER (e.g., the NOAA CPC precipitation dataset and the NCEP/NCAR global 2.5 degree products) are updated in near real-time, and our strategy is to update COASTER with the additional data several times a year. To simplify the process of updating datasets on COASTER, we have developed a collection of tools and procedures for rapidly integrating new data, including approaches for 1) preprocessing (e.g., converting file types and/or extracting single date rasters from multi-banded datasets); 2) producing new projection translation functions (*i.e.*, for converting the latitude and longitude values entered on the web-based interface into map units); 3) modifying the details stored in the dataset SQL database to reflect the new temporal extent of the updated dataset; and 4) modifying the COASTER interface to reflect the presence of the expanded years available.

As a beta system, several challenges remain unsolved for the COASTER system. One such issue is the automated production of metadata that meets accepted standards. Since the COASTER system is independent of the

data creation process, we are limited to providing only the metadata associated with the raw datasets that was generated by the producers of those datasets. Our preliminary method of addressing this issue is to supply a small text file describing the user-selected summarization options (available on the FTP link along with the .tif file) with all COASTER outputs. For now this is an acknowledged limitation of COASTER that we plan to address as resources allow.

The ultimate value of COASTER can only be assessed once the system is fully operational and more details are collected to document and assess system usage and performance. From the outset the COASTER system was designed with long-term viability in mind. As such, we have strived to make a system that overcomes the very real challenges of cost, processing and delivery time for output, scalability, ease of updating or adding new datasets, system robustness, and applicability for use beyond the research community. A major element of this long-term thinking is keeping COASTER as small and simple as possible while still fulfilling its mission of providing needed data products. The rationale behind this decision was our desire to create a system that was inexpensive to operate and not dependent on tools such as those provided within an ArcGIS or Google Maps environment (*i.e.*, potentially necessitating greater maintenance costs in response to changing protocols in software on which COASTER depends). The trade-off of the decision to keep COASTER simple, however, is that COASTER lacks the data visualization tools of more ambitious projects like the online Water Data Discovery and Retrieval system [13]. Depending on user feedback and funding, we may attempt to enhance COASTER by integrating output datasets into a web mapping system that enables users to interact with data using only a web browser. A likely platform for this functionality is ClimateScape (www.ClimateScape.net), another online tool funded by NSF and developed by YERC and its sister organization HyPerspectives, Inc. While the full suite of COASTER functions are unlikely to transition into ClimateScape, a wide variety of data summaries (e.g., daily average conditions) could be made in advance and stored for use within ClimateScape as needed.

6. Conclusion

COASTER represents an important advance among systems that distribute environmental raster datasets due to its ability to quickly and easily produce customized summarizations from high temporal resolution data. The functionality and data made freely available by COASTER significantly lower the barrier of entry for those working with datasets such as daily climate data by enabling researchers with limited GIS or remote sensing expertise to create data tailored to their needs. The structure of the COASTER system is flexible enough to accommodate many types of environmental raster datasets with relative ease. The highly scalable design of the COASTER system increases the likelihood that it will persist into the future (at a relatively low cost) while preserving the core functionality that makes it useful to researchers. In total, COASTER has the potential to benefit a wide range of users for years to come by greatly reducing the effort required to create datasets needed for research and resource management.

7. Acknowledgements

Partial funding for the COASTER project was provided by 1) NASA Ecological Forecasting (RRSC-SHC) NNX-08AM70G; 2) NASA Biological Response to Climate (A.30 BioClim) NNX11AK90G; 3) NSF Field Station and Marine Lab (FSML) 0829495; and 4) NSF EPSCoR Track 1 Montana University System System PG12-67600-01. Special thanks to the individuals and organizations responsible for creating and making available the datasets hosted on COASTER, including 1) Jared Oyler at the University of Montana; 2) Rama Nemani, Andy Michaelis, and Forrest Melton at the NASA Ames Research Center; and 3) all the scientists contributing to datasets hosted by the National Oceanic and Atmospheric Administration Earth System Research Laboratory, Physical Sciences Division.

REFERENCES

[1] R. Nemani, H. Hashimoto, P. Votava, F. Melton, W. Wang, A. Michaelis, L. Mutch, C. Milesi, S. Hiatt and M. White, "Monitoring and Forecasting Ecosystem Dynamics Using the Terrestrial Observation and Prediction System (TOPS)," *Remote Sensing of Environment*, Vol. 113, No. 7, 2009, pp. 1497-1509. doi:10.1016/j.rse.2008.06.017

[2] R. Nemani, , P. Votava, A. Michaelis, M. White, F. Melton, J. Coughlan, K. Golden, H. Hashimoto, K. Ichii, L. Johnson, M. Jolly, C. Milesi, R. Myneni, L. Pierce, S. Running, C. Tague and W. Pang, "Terrestrial Observation and Prediction System (TOPS): Developing Ecological Nowcasts and Forecasts by Integrating Surface, Satellite and Climate Data with Simulation Models," The Research and Economic Applications of Remote Sensing Data Products, American Geophysical Union, 2005. https://c3.nasa.gov/nex/resources/1/ https://c3.nasa.gov/nex/static/media/publication/TOPS_AGU.pdf

[3] P. E. Thornton, S. W. Running and M. A. White, "Generating Surfaces of Daily Meteorological Variables over Large Regions of Complex Terrain," *Journal of Hydrology*, Vol. 190, 1997, pp. 214-251. doi:10.1016/S0022-1694(96)03128-9

[4] C. Daly, R. P. Neilson and D. L. Phillips, "A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain," *Journal of Applied*

Meteorology, Vol. 33, No. 2, 1994, pp. 140-158. doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2

[5] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne and D. Joseph, "The NCEP/NCAR 40-Year Reanalysis Project," *Bulletin of the American Meteorological Society*, Vol. 77, No. 3, 1996, pp. 437-471. doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

[6] R. W. Higgins, J. E. Janowiack and Y.-P. Yao, "A Gridded Hourly Precipitation Data Base for the United States (1963-1993)," NCEP/Climate Prediction Center Atlas 1, National Centers for Environmental Prediction, 1996, 46 pp.

[7] F. Mesinger, G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish and W. Shi, "North American Regional Reanalysis," *Bulletin of the American Meteorological Society*, Vol. 87, No. 3, 2006, pp. 343-360. doi:10.1175/BAMS-87-3-343

[8] T. Wang, A. Hamann, D. L. Spittlehouse and S. N. Aitken, "Development of Scale-Free Climate Data for Western Canada for Use in Resource Management," *International Journal of Climatology*, Vol. 26, No. 3, 2006, pp. 383-397. doi:10.1002/joc.1247

[9] M. S. Mbogga, A. Hamann and T. Wang, "Historical and Projected Climate Data for Natural Resource Management in Western Canada," *Agricultural and Forest Meteorology*, Vol. 149, No. 5, 2009, pp. 881-890. doi:10.1016/j.agrformet.2008.11.009

[10] T. D. Mitchell and P. D. Jones, "An Improved Method of Constructing a Database of Monthly Climate Observations and Associated High-Resolution Grids," *International Journal of Climatology*, Vol. 25, No. 6, 2005, pp. 693-712. doi:10.1002/joc.1181

[11] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones and A. Jarvis, "Very High Resolution Interpolated Climate Surfaces for Global Land Areas," *International Journal of Climatology*, Vol. 25, No. 15, 2005, pp. 1965-1978. doi:10.1002/joc.1276

[12] D. J. Kriticos, B. L. Webber, A. Leriche, N. Ota, I. Macadam, J. Bathols and J. K. Scott, "CliMond: Global High-Resolution Historical and Future Scenario Climate Surfaces for Bioclimatic Modelling," *Methods in Ecology and Evolution*, Vol. 3, No. 1, 2012, pp. 53-64. doi:10.1111/j.2041-210X.2011.00134.x

[13] M. Huang, D. R. Maidment and Y. Tian, "Using SOA and RIAs for Water Data Discovery and Retrieval," *Environmental Modelling & Software*, Vol. 26, No. 11, 2011, pp. 1309-1324. doi:10.1016/j.envsoft.2011.05.008

