Climate Interpolation and Cross Validation Toolkit in the Reynolds Creek Watershed

Tucker Chapman, Donna Delparte Ph.D. - Idaho State University, Department of Geosciences

Idaho State UNIVERSITY

discover OPPORTUNITY

Introduction

Current methods for producing spatially distributed raster grids for snow models are time consuming and often require researcher input leading to possible biased results. Empirical Bayesian kriging (EBK) is a novel approach which automates the estimation of the semi-variogram model used for kriging (Krivoruchko, 2012). Raster grids are spatially distributed rasters of various parameters which are created using point data collected at weather stations. The Image SNOWcover and mass BALance (ISNOBAL) model is used to predict seasonal snowmelt (Marks, et al., 1999, Susong et al. 1999). Using Esri's ArcGIS tools an automated approach has been developed to create raster grids for snow modeling in ISNOBAL. Kriging is a commonly used geostatistical approach for interpolating known point values to unknown points. In order to compare methods/approaches for generating raster grids, a cross validation of EBK was conducted. Results reveal an overall low mean average error (MAE) and Root Mean Square Error (RMSE) for most of the parameters. Cross validation was performed using a leave-one-out method for the 2014 water years of Johnston Draw and the 2008 water year of Reynolds Creek. The output from these automated tools can be used to force the ISNOBAL model. Simple graphical user interfaces have been developed to give researchers an easy to use tool for producing raster grids. These tools can also be used for any other model that requires a raster grid as input.

Objective 1

Climate interpolation toolkit. Creates spatially distributed estimates of climate parameters

Objective 2

Climate interpolation cross validation tool. Comparison of interpolation method accuracy.

Reynolds Creek South Watershed

The Reynolds Creek South watershed (RCS) in southwest Idaho is an experimental watershed with data collected since 1962 from over 30 stations of varying operation, duration, and type. The concentration of recording stations in the RCS watershed makes it ideal for building climate interpolation tools (Slaughter, et al., 2001).

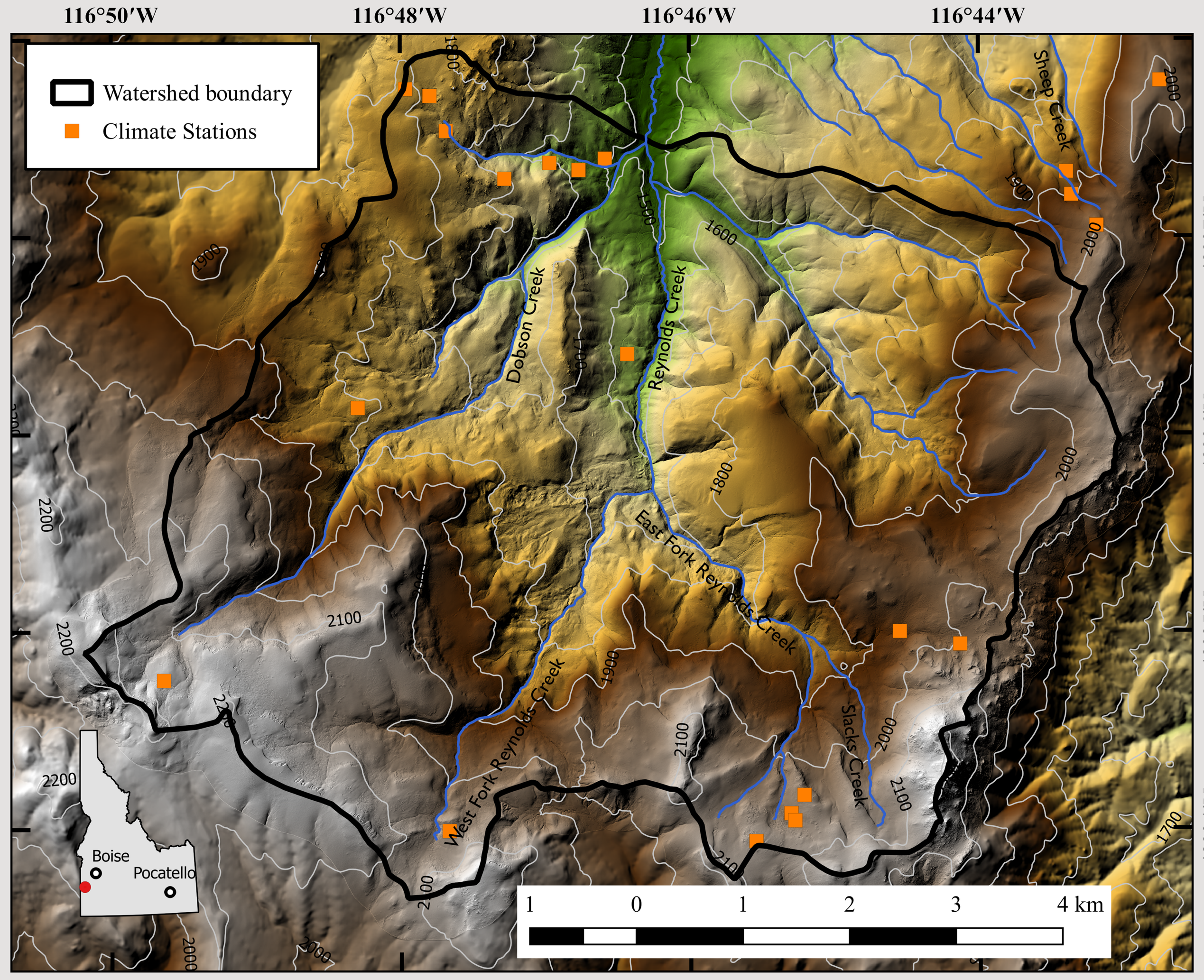

Figure 1: Reynolds Creek South Watershed boundary. Vegetation in high elevations are alpine and forest type. Lower elevations consist of typical Great Basin Desert plants. Precipitation ranges from ~230 mm to > 1100 mm for low to high elevations. Soils are granitic, volcanic and lake sediments (Slaughter, et al., 2001).

Methods

Residuals are determined using a leave-one-out method of cross validation. A python script has been written to automate the process of leave-one-out. For each time step (1-6 hour steps) several grids are interpolated leaving out one station at a time. The predicted value for the left out station is recorded as well as the observed value. The mean average error (MAE) and root mean square error (RMSE) can then be calculated using the following formulas. Graphs are produced using the python module matplotlib (Hunter, 2007).
MAE is the absolute value of the difference between the modeled value and the observed value. Then calculating the mean of all values. RMSE is calculated by squaring the difference between the modeled value and the observed value. The average is calculated then square rooted. Using these two measures we can determine general model error.

Root Mean Squared Error

$$\sqrt{\frac{\Sigma (Obs - M)^2}{Count}}$$

Mean Absolute Error

$$\frac{\sum |(Obs - M)|}{Count}$$

Interpolated Surface - Air Temperature

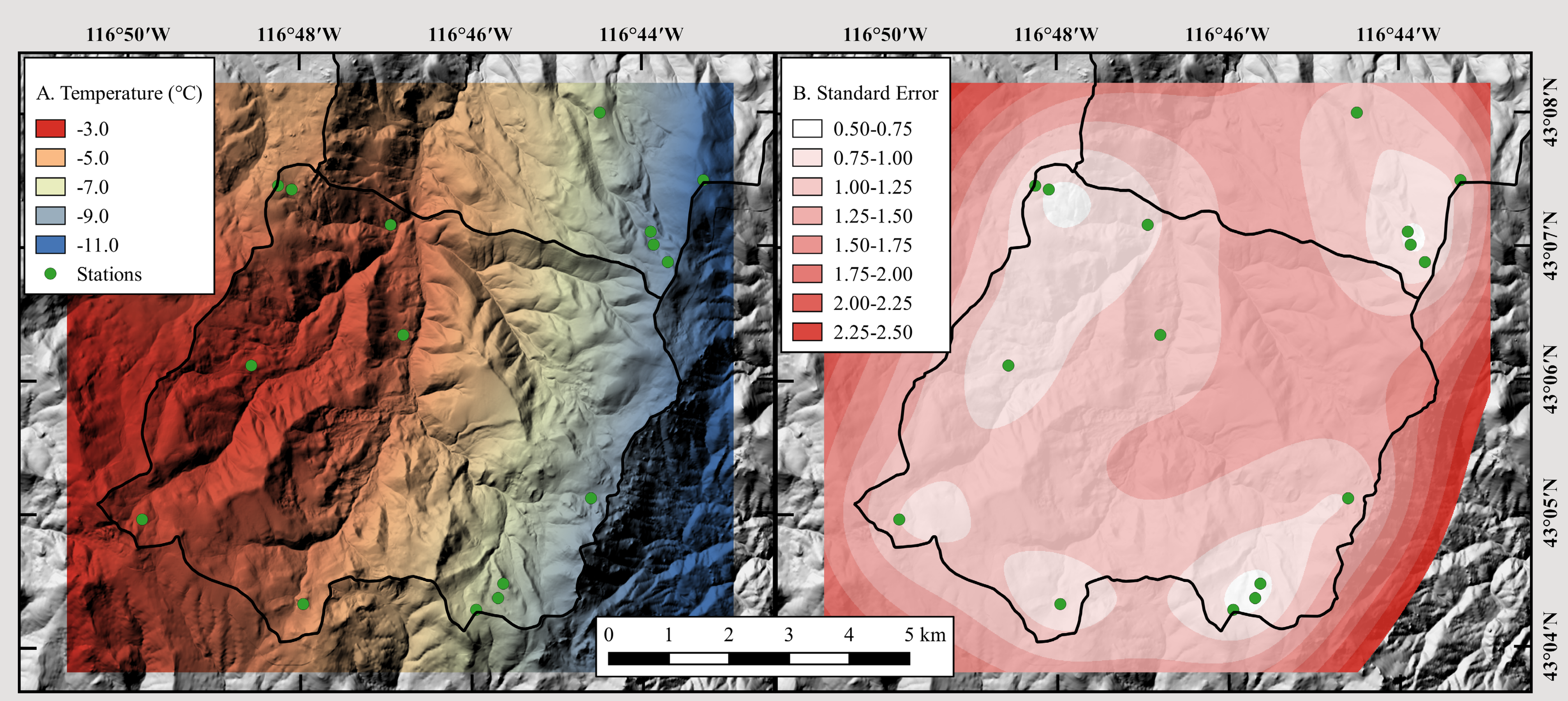

Figure 2: An example air temperature raster surface created using the interpolation tools. Several of these are created in order to calculate the MAE and RMSE. Data for this map is from the 2008-01-01 12:00 - 13:00 time step.

Results

Example graphs from the cross validation toolkit. Provides a quick visualization check of outliers and problem stations.

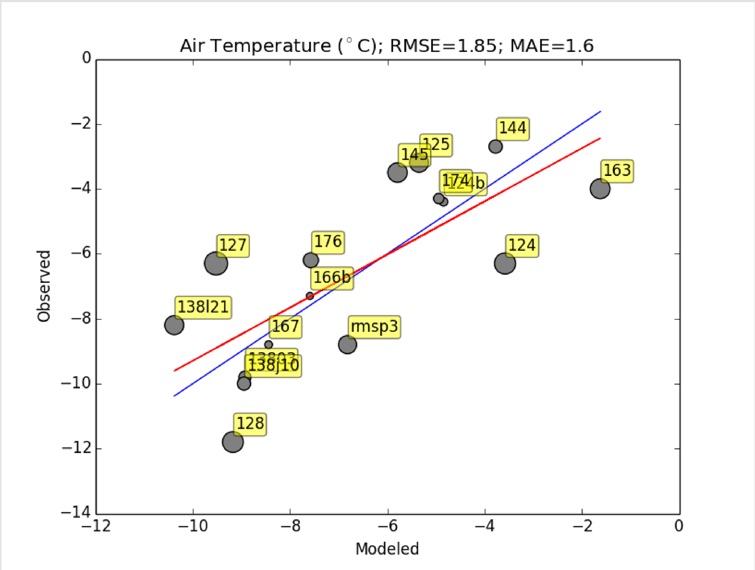

Figure 3: Air temperature output graph from the cross validation toolkit for January 1, 2008 12:00-13:00 time step in RCS watershed. The closer the red line is to the blue line the better fit the model is. Labels indicate the station names. This graph is an example of a well balanced cross validation run.

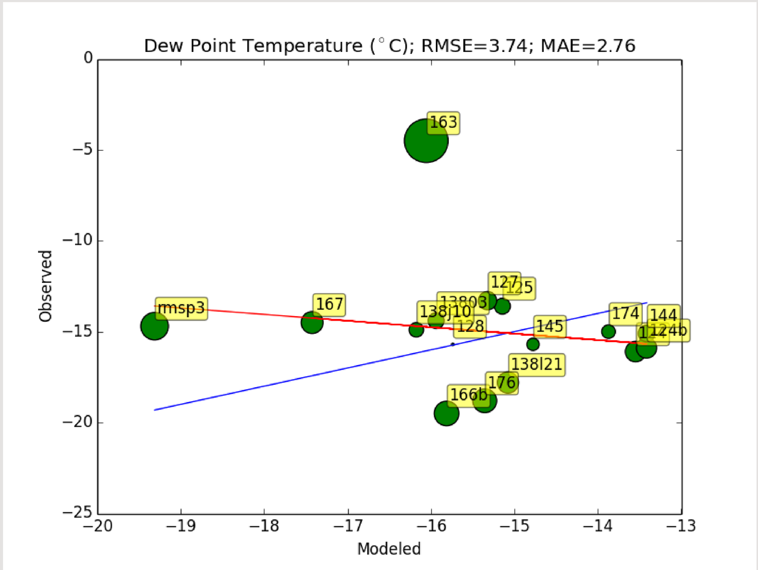

Figure 4: Dew point temperature output graph from the cross validation toolkit for the same time step. The graphs from the toolkit can be used to find outliers and other QAQC problems that can be useful for researchers. Station 163 is in the southwest corner of the watershed with no other nearby stations.

Data Tables

Data tables providing daily averages of root mean square and mean absolute errors for both the Johnston Draw and Reynolds Creek South watersheds. Also provided are sample times for the time needed to create one raster surface of each parameter type.

Johnston Draw

2013-10-01 to 2013-10-02			
Parameter	MAE	RMSE	Time(s)
Air Temperature (°C)	0.42	0.59	6.30
Dew Point (°C)	0.47	0.76	6.62
Vapor Pressure (Pa)	23.40	30.98	6.71

2014-01-01 to 2014-01-02			
Parameter	MAE	RMSE	Time(s)
Air Temperature (°C)	0.45	0.62	8.04
Dew Point (°C)	0.24	0.30	8.67
Vapor Pressure (Pa)	13.69	17.09	8.42

2014-03-01 to 2014-03-02			
Parameter	MAE	RMSE	Time(s)
Air Temperature (°C)	0.21	0.26	8.33
Dew Point (°C)	0.20	0.24	9.70
Vapor Pressure (Pa)	11.37	13.93	9.29

MAE - Mean absolute error
RMSE - Root mean squared error
Time - Average time to create each raster grid

Reynolds Creek

2008-01-01 to 2008-01-02			
Parameter	MAE	RMSE	Time(s)
Air Temperature (°C)	1.38	1.68	42.52
Dew Point (°C)	1.83	2.59	44.63
Vapor Pressure (Pa)	28.93	48.58	43.99
Snow Depth (cm)	16.19	21.45	53.32

2008-01-01 to 2008-01-02			
Parameter	MAE	RMSE	Time(s)
Air Temperature (°C)	1.23	0.92	40.40
Dew Point (°C)	0.92	0.71	41.43
Vapor Pressure (Pa)	31.92	24.67	40.84
Snow Depth (cm)	44.01	33.62	55.09

Precipitation (mm)			
Parameter	MAE	RMSE	Time(s)
2008-12-24 to 2008-12-25	1.20	0.87	38.56
2009-06-04 14:00 - 17:00	0.75	0.46	41.36

Discussion

The climate interpolation and cross validation toolkits used in this study can be used to produce the spatially distributed input grids required to run ISNOBAL and other spatial distributed climate models. Cross validation shows that the empirical Bayesian kriging interpolation method provides accurate results for climate parameters for the Reynolds Creek South watershed. Vapor pressure and Snow depth have the highest interpolation error of the five parameters tested. These high error values are likely due to the larger ranges in value between stations and the higher variability of the parameter type. Air temperature, dew point temperature, and precipitation all have very low errors.
A significant part of preparing for any virtual watershed models is the time required to prepare the climate grids to force the model. The interpolation toolkit is completely automated making it easy to produce new data to run ISNOBAL (or other distributed climate models such as PRMS (Leavesley, et al., 1983) or HydroGeoSphere (Therrien, et al., 2010)). The cross validation toolkit produces several interpolation surfaces for each time step and records the amount of time required to calculate the surface. Several averages are included in the above data tables. The smaller Johnston Draw sub-watershed requires ~9 seconds to produce a single grid. The Reynolds Creek South watershed requires ~45 seconds. Over 8000 grids are required to run a full ISNOBAL so in order to produce a full dataset for Johnston Draw it would require about one full day of computation time.

Future Research

The cross validation toolkit will be used to verify that the interpolation methods used to produce model input grids are accurate. The cross validation toolkit will be used to compare these different parameters:

- Empirical Bayesian Kriging models
- Different DEM resolutions (90m, 30m, 10m)
- Comparison of empirical Bayesian kriging to other interpolation methods

IDAHO EPSCoR

This material is based in part upon work supported by: The National Science Foundation under grant number #IIA-1329513 Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing In Science & Engineering, 9(3), 90–95.
Krivoruchko, B. K., & Team, S. D. (2012). Empirical Bayesian Kriging Implemented in ArcGIS Geostatistical Analyst, 6–10.
Lichty, R. W., B. M. Thoutman, and L. G. Saindon. Precipitation-runoff modeling system: User's manual. Colorado, CO: US Geological Survey, 1983.
Marks, D., Domingo, J., Susong, D., Link, T., & Garen, D. (1999). A spatially distributed energy balance snowmelt model for application in mountain basins. Hydrological Processes, 13(12-13), 1935–1959.
Slaughter, C. W., Marks, D., Flerchinger, G. N., Vactor, S. S. Van, & Burgess, M. (2001). Thirty-five years of research data collection at the Reynolds Creek Experimental Watershed , Idaho , United States, 37(11), 2819–2823.
Susong, D., Marks, D., & Garen, D. (1999). Methods for developing time-series climate surfaces to drive topographically distributed energy- and water-balance models. Hydrological Processes, 13(May 1998), 2003–2021.
Therrien, R., et al. "HydroGeoSphere: a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport." Groundwater Simulations Group, University of Waterloo, Waterloo, ON (2010).

Contact Information

Tucker Chapman
chaptuck@isu.edu
Graduate Research Assistant
Idaho State University
Pocatello, ID 83201

Donna Delparte Ph.D.
delparte@isu.edu
Assitance Professor
Idaho State University
Pocatello, ID 83201