Environment and organisms - Task #491

Methods comparison

09/11/2012 05:21 AM - Benoit Parmentier

Status:	In Progress	**Start date:**	09/11/2012
Priority:	Normal	**Due date:**	
Assignee:	Benoit Parmentier	**% Done:**	0%
Category:	Climate	**Estimated time:**	0.00 hour
Target version:			
Activity type:	Coding/analysis		

Description

Decision on the choice of the best interpolation method to use in the upcoming product necessitates a systematic comparison of interpolation methods. This task relies mainly on codes and results developed during the interpolation stage for the Oregon case study over a full year (365 dates).

History

#1 - 09/13/2012 02:39 PM - Benoit Parmentier

For now, interpolation methods are assessed using:

1) Visual patterns from prediction

This is a visual assessment of the outputs to ensure that the predicted values are sensible in terms of the spatial configuration of the study area. This may need to be translated in some metrics at a later stage.

2) Accuracy metrics over 365 dates (one year)

- MAE, RMSE, R2 for training and testing using box plots and averages
- Additional metrics are available depending on the interpolation methods (i.e. AIC for GAM)
See the following script: 170edade

3) Using multi sampling

This consists in changing the training and testing stations using variable proportions and random sampling.
Proportions of hold out are varied from 10 to 70% by step of 10% and testing stations are randomly selected 15 times for each proportion. The current script allows for changing the portions, step and number of random samples (i.e. replication).
See the following scripts: 4719fdd7 ; e7bf2d1b ; 101f27b0 ; 69864891.

4) Accuracy and spatial distance to closest training station

It is expected that average accuracy decreases as one moves away from training stations. Plots are created by calculating the average accuracy as a function of distance to closest training station. Currently the average MAE is calculated for 15 bins centered from 5 to 135km. Average of accuracy metrics are then assessed for different methods and interpolated models.
See the following script: 2bdb1ff5

#2 - 10/19/2012 11:01 AM - Benoit Parmentier

- File IPLANT_working_meeting_10182012_Benoit_update.pdf added

This is a first summary of the methods comparison for the interpolation of maximum temperature (tmax) in Oregon. More updates will come in term of coding. Five methods are compared: Kriging, GWR, GAM, CAI and Fusion. I used six procedures for accuracy assessment: 1) accuracy metrics, 2) multisampling, 3) accuracy in term of closest training station, 4) map visualization 5) accuracy in term of density of stations, 6) accuracy profiles at specific stations.

#3 - 12/12/2012 04:18 PM - Benoit Parmentier

- File IPLANT_update_additional_analysis_part_I_11042012.pdf added

This is the first part of the additional analyses that was carried out following the five methods comparison. This presentation was shown early November during the IPLANT meeting.
It includes:
1) comparison of accuracy with results using all stations at the monthly time scale
2) screening of LST and ELEV_SRTM with a quick look at LST monthly and TMax averages over year 2010
3) simplified models for CAI and comparison to fusion with Kriging

Codes relevant to the analyses can be found in the repository:

-production of boxplots to compare the five methods and maps of predicted surface: f82d4df1
-transects,accuracy assessment in term of distance to closest training station and multi-sampling: 5e7d95a7
-checking of extremes values in LST and elevation, exploratory analyses: c352e9e1

#4 - 12/12/2012 04:36 PM - Benoit Parmentier

- File IPLANT_update_additional_analysis_part_II_12072012.pdf added

This is the second part of the additional analyses on method comparison.This presentation focuses on the analyses of residuals at GHCN station locations and the comparison of CAI and fusion interpolated surfaces.
The presentation includes:
1) detailed plots to compare fusion (kriging) and CAI (Kriging) for two dates: 20100103 and 20100901
2) Study of differences between CAI and fusion predictions using the interpolated surface and covariates (LST and ELEV_SRTM)
3) Transects through the predicted surfaces in relation to elevation (ELEV_SRTM).

Codes related to the analyses can be found here:
residuals analyses update on plots of residuals per elevation classes: d07e9853
residuals analyses, spatial transects through stations with difference and elevation: 30f84063

Files

IPLANT_working_meeting_10182012_Benoit_update.pdf	4.08 MB	10/19/2012	Benoit Parmentier
IPLANT_update_additional_analysis_part_I_11042012.pdf	2.57 MB	12/13/2012	Benoit Parmentier
IPLANT_update_additional_analysis_part_II_12072012.pdf	3.84 MB	12/13/2012	Benoit Parmentier