Environment and organisms - Task #364

Integrate spatial variables and structure in the GAM methodology

02/23/2012 02:24 PM - Benoit Parmentier

Status:	New		**Start date:**	02/23/2012
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:				
Activity type:				

Description
MODIS LST time series provide information on the structure and spatial variability of skin temperature. This information may be used to improve prediction from GAM by incorporating spatial variables and structure in GAM.

History

#1 - 03/11/2012 11:54 AM - Benoit Parmentier

One of the initial step in the methodology will be to reproduce the GWR model from PRISM. I am experimenting with GWR to try to produce prediction using one date.

#2 - 03/19/2012 03:41 PM - Benoit Parmentier

GWR predictions were produced using the sgwr package in R.

The following specifications were used to run the models:

Dependent variable: tmax
Independent variables: lon, lat, ELEV_SRTM, Eastness, Northness, DISTOC
Bandwidth: determined from the data by CV (one leave out approach).
Weight function model: Gaussian
proportion of hold out: 0 , 30, 50%, 70%
validation: RMSE fit

On Thursday 03/15, Jim Regetz, Brian McGill and Benoit Parmentier had a discussion on the best approach to validate GWR models. Two validation approaches have emerged from the discussion:

1) Approach 1
First GWR is performed on the training dataset to produce coefficients at every training stations. Second a surface of parameters (slope coefficient) is obtained by interpolation (Kriging). Third, tmax values at testing samples are then obtained by applying the parameters at the testing locations. Fourth an RMSE is calculated for the testing dataset.
2) Approach 2
First, GWR is performed on the training dataset and the bandwidth is obtained. Second, the training bandwidth is then used when running GWR on the testing dataset. Third, coefficients produced at testing sites are used to predict tmax values for testing samples. Fourth an RMSE is calculated for the testing dataset.

Early results indicate that RMSE values on average higher for the GWR models compared to earlier models using GAM. More test may be needed to check the dependency of the results on the sample and dates chosen. A powerpoint will be created to summarize some of the results.

Some References:
Harris P., A.S. Fotheringham, R. Crespo, M. Charlton. (2010). The Use of Geographically Weighted Regression for Spatial Prediction: An Evaluation of Models Using Simulated Data Sets. Math Geosci:: 657–680

Llyod C.D. (2010). Nonstationary models for exploring and mapping monthly precipitation in the United Kingdom. INTERNATIONAL JOURNAL OF CLIMATOLOGY Int. J. Climatol. 30: 390–405.

Wimberly1 M.C., M. J. Yabsley, A. D. Baer1, V. G. Dugan, and W. R. Davidson (2008). Spatial heterogeneity of climate and land-cover constraints on distributions of tick-borne pathogens land-cover constraints on distributions of tick-borne pathogens Global Ecology and Biogeography, (Global Ecol. Biogeogr.) 17, 189–202.

#3 - 05/01/2012 02:08 PM - Benoit Parmentier

Kriging has been done so far by using variograms fitted from stations' locations. The next step is to examine the fitting of variogram directly from LST images to capture its spatial structure.

#4 - 05/09/2012 03:32 PM - Benoit Parmentier

I am currently exploring a two stage regression involving:
1) Step 1: GAM models using Lat, long, ELEV_SRTM, DISTOC, ASPECT, Land Cover, Monthly LST
2) Step 2: Kriging on models from step 1

The code is currently being written in R. I expect to be done this week or next week.

#5 - 06/05/2012 12:27 AM - Benoit Parmentier

Two stage regression: the GAM+Kriging script was updated to integrate specific and general diagnostics:
general diagnostics: RMSE, MAE,ME (BIAS), R2 (general for all methods)
specific diagnostics: AIC,DEV,GCV,RMSE_f   (specific to GAM)

Code modifications, 328689cf

This modification will allow us to compare results to the literature. Results from running the code for 10 dates show that the MAE is less affected by the skewness of the residuals as expected. The kriging code must still be modified to fit automatically variograms.