Phenological modelling using volunteered observations and machine learning methods

***Raul Zurita-Milla**, Emma Izquierdo-Verdiguier, Hamed Mehdipoor*

Department of Geo-Information Processing, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede (The Netherlands)

Phenological models predict the timing of recurrent biological events like the appearance of the first leaf in deciduous plant species. Since in many ecosystems these timings are driven by environmental conditions, phenology has become a popular indicator for climate change. In addition, having reliable and robust phenological models is important for several applications like nature conservation, ecology, agriculture and even public health (e.g. hay fever).

In this study we use volunteered phenological observations curated and/or collected by the USA National Phenological Network to develop several machine learning-based phenological models. In particular we use long term records of first leaf and first bloom for lilac and honeysuckle species. The required explanatory variables for these models (i.e. the regressors) were extracted from DAYMET, a high spatial resolution (1 km^2) daily gridded dataset of weather variables for the period 1980 to the present.

Different non-linear machine learning methods were tested, from classical ones based on artificial neural networks or decision trees to more recent ones like Gaussian process or Kernel-regression methods. For this we used a convenient regression toolbox LÁZARO-GREDILLA et al. (2013). All models were intercompared and evaluated according to several quantitative measures of accuracy (systematic and unsystematic RMSE and MAE), bias (ME) and goodness-of-fit (with the Pearson's correlation coefficient) and benchmarked against standard phenological models (e.g. based on linear regression).

Results indicate that non-linear methods are able to accurately capture the dynamics present in the volunteered phenological observations. Further work includes testing the top models using other plant species and the identification of robust (different environments) and meaningful (biologically) regressors as well as on the application of the top models at a continental scale and for the complete period covered by the DAYMET dataset.

References:
LÁZARO-GREDILLA, M., TITSIAS, M.K., VERRELST, J., CAMPS-VALLS, G., 2014: Retrieval of Biophysical Parameters with Heteroscedastic Gaussian Processes. *IEEE Geoscience and Remote Sensing Letters*, **11(4)**, 838-842.

Keywords: modelling, machine learning, phenological volunteered observations