Spatial and statistical modelling of phenological data using 'R'

Daniel Doktor*
Imperial College

This paper analyses the spatio-temporal patterns of Land Surface Phenology (LSP) in Germany. LSP is the study of the spatio-temporal development of vegetated land surface as revealed by synoptic sensors, whether space borne remote sensors or in situ observational networks. LSP provides a critical window on the local consequences of global change.

The phenological data used for this study come from the phenological network of the German Weather Service and is managed by the relational database system 'Oracle'. The packages **'DBI'** and **'ROracle'** (64 bit application) were utilized for communicating and interfacing between 'R' and the database. The package **'gstat'** was applied for geostatistical modelling purposes such as Variogram estimation and Kriging. Spatial interpolation of phenological ground observations was carried out either using Detrended Kriging (referring to average elevation gradients) or applying External Drift Kriging. Both interpolation methods performed on a similar level with estimation errors between 3-9 days (using cross-validation). The results (d_{BB}) were visualised using the packages **'lattice'** and **'maptools'**.

The interpolated budburst dates from ground observations were compared to satellite derived dates of green-up (d_{GU}). Mean, modus and median of the frequency distribution of $d_{BB} - d_{GU}$ indicate that satellite derived green-up preceded observed tree budburst dates on average by 3 days.

In order to quantify the influence of cold spells on the temporal evolution (pace) of spring time phenology we used Gaussian Mixture Modelling. Using this methodology it was possible to quantitatively characterise the frequency distributions of observed budburst dates. Mixture components could be identified either via Expectation-Maximisation (EM) or via an optimisation algorithm. The EM algorithm was initialised by hierarchical clustering (package **'mclust'**) for parameterised Gaussian Mixture Models. The number of clusters and the clustering model is chosen to maximise the Bayesian Information Criterion (BIC). Secondly, an optimisation algorithm (**'base'**: optim) was applied on the minimisation of several (maximum four) Gaussian Mixture Functions. Based on Akaike's Information criterion it was decided which optimised function was the most appropriate. The identified mixture components also formed the methodological base for a new outlier detection algorithm to be applied within huge phenological databases.

Additionally, space-time correlations of the phenological ground observations were analysed. Every phenological station was compared to all others provided the respective station pair had at least 20 identical years of observations. Dependent on the geographical distance of the pair correlation coefficients were assigned to certain distance categories for each station. Consequently, it was possible to determine how phenological time series correlate over space and which stations showed noticeably different trends when compared to the Grand Mean.

It is the authors intention to include phenologically related functionalities into the 'R'-package **pheno**.

*Correspondence: Daniel Doktor
d.doktor@imperial.ac.uk
Prince Consort Road, RSM Building
SW7 2BP London