Spatial and statistical modelling of phenological data using ‘R’

- Are ground observations comparable to satellite data?
- Can we characterise the pace of phenological development in spring time?
- How far do temporal correlations of phenology extend geographically?

Contact: d.doktor@imperial.ac.uk

What is phenology?

- Phenology is the study of the timing of recurring life cycle events
- For plants such events are budburst, leaf unfolding, blossoming, fruit ripening, leaf colouring etc.

Satellite and ground observational data

- Dynamic filtering BISE (Viovy et al. 1992)

Spatial interpolation of ground observations

- External Drift Kriging (EDK), thereby incorporating elevation as secondary information
- Detrended Kriging based on a Global elevation gradient:
  \[ g_w = \frac{\sum (\frac{\Delta f}{\Delta h}) \cdot w_i}{\sum w_i} \]
- Used package for geostatistical purposes: gstat

Phenological observation stations in Germany

Station data format: X/Y, elevation, budburst date
Oracle database, interfacing and communicating with packages ROracle and DBI

Computed budburst date based on NDVI values

Apple flowering
Ash budburst
Beech leaf colouring

Federal States
Observation station

Dynamic filtering BISE (Viovy et al. 1992)
Crossvalidation of interpolation methods

- Both interpolation techniques are of nearly the same quality (green=EDK, red=Detrended Kriging)
- The mean MAE is about 5 days for each species and method

EDK vs NDVI

Selected NDVI-pixels for comparison buffered by deciduous forest according to CORINE

⇒ The obtained subset contains only a fraction of the original information (~0.02 %)

Results:
- Mean difference of 3.3 days for 1989-1997
- Satellite derived green-up preceded ground observations
- Average correlation coefficient $r = 0.38$

Problem:
- Heterogeneity in vegetation cover affects NDVI-signal

Gaussian Mixture Models

$ f_m(x) = p_1 f_1(x) + \ldots + p_n f_n(x) $  
$p_1, \ldots, p_m$ positive numbers summing to one

$f_1(x), \ldots, f_n(x)$ the component densities

Different modelling approaches:
- Optimisation algorithm  => ‘base’
  Akaike’s Information Criterion (AIC) based on chi-square at 0.05 significance level
- Clustering via EM-algorithm  => package ‘mclust’
  Bayesian Information Criterion (BIC)

Optimisation algorithm

Frequency distributions of observations of Oak (green line) modelled using between 1-4 Mixture distributions

<table>
<thead>
<tr>
<th>Year</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\mu_3$</th>
<th>$\mu_4$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\sigma_3$</th>
<th>$\omega_1$</th>
<th>$\omega_2$</th>
<th>$\omega_3$</th>
<th>$\omega_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>134.95</td>
<td>82.78</td>
<td>116.86</td>
<td>137.51</td>
<td>4.78</td>
<td>11.34</td>
<td>3.83</td>
<td>6.78</td>
<td>0.15</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>1980</td>
<td>132.09</td>
<td>135.33</td>
<td>116.31</td>
<td>131.55</td>
<td>4.70</td>
<td>11.34</td>
<td>3.83</td>
<td>6.78</td>
<td>0.15</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>1981</td>
<td>104.54</td>
<td>116.31</td>
<td>131.55</td>
<td>137.51</td>
<td>4.70</td>
<td>11.34</td>
<td>3.83</td>
<td>6.78</td>
<td>0.15</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>1982</td>
<td>128.15</td>
<td>135.14</td>
<td>106.38</td>
<td>3.78</td>
<td>6.78</td>
<td>0.15</td>
<td>0.37</td>
<td>0.32</td>
<td>0.16</td>
<td>0.62</td>
<td>0.38</td>
</tr>
</tbody>
</table>

$\text{orange} = 1 \text{ component(s)}$
$\text{blue} = 2$
$\text{red} = 3$
$\text{black} = 4$
**EM algorithm**

Differences in number of detected Gaussian Mixtures and their characterising values.

Permutation of the order and changing of initial values had no effect on the outcome of EM-algorithm.

---

**Space-time correlations of ground observations**

- Time series’ correlation of station pairs is assigned to distance categories.

<table>
<thead>
<tr>
<th>Distance category</th>
<th>0-25</th>
<th>25-50</th>
<th>50-100</th>
<th>100-200</th>
<th>200-300</th>
<th>300-400</th>
<th>400-500</th>
<th>500-600</th>
<th>600-700</th>
<th>700-800</th>
<th>800-900</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Years are only chosen when both stations have observed.

**Results:**
- Low correlations for stations with at least 50 observed years $r=0.45$.
- For years with a unimodal distributions $r=0.65$.

**Results:**

after Koenig et al. (1998)

---

**Detecting stations with reversed trends**

- Focus on single stations with negative correlations over all distance categories.
- Detection of reversed trends when their annual observations are compared to the Grand Mean.

**Conclusions**

Solid interpolation methodologies allow the comparison of ground and satellite observations. Due to heterogeneity of ground vegetation correlations between the two are weaker than expected.

The pace of phenological development can be characterised quantitatively using Gaussian Mixtures. Between 1-4 mixtures could be identified reflecting strongly variable weather patterns during springtime.

Temporal correlations of phenological data extend over relatively large distances. The correlation’s magnitude depends on weather patterns experienced within each analysed year.

**Origin of reversed trends?**

- Change in immediate environment of observed species (microclimate).
- Falsely recorded phenological phases.