Spatial and statistical modelling of phenological data using 'R'

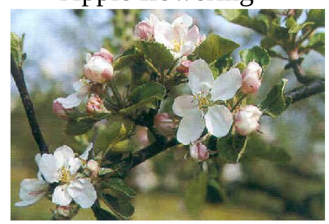
- Are ground observations comparable to satellite data ?
- Can we characterise the pace of phenological development in spring time ?
- How far do temporal correlations of phenology extend geographically ?

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.

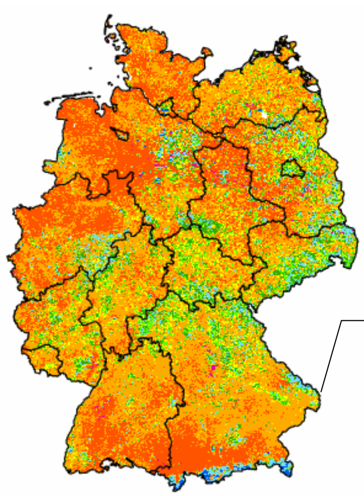
Apple flowering



Ash budburst

Beech leaf colouring

Satellite and ground observational data

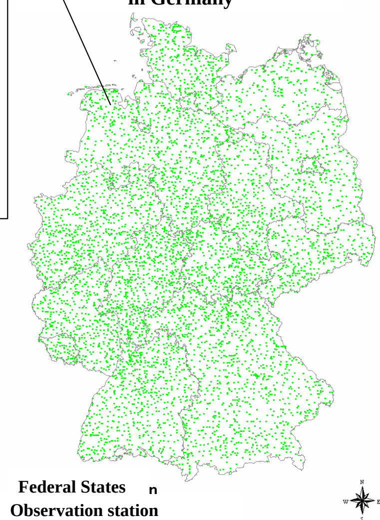


Station data format: X/Y, elevation, budburst date
Oracle database, interfacing and communicating with packages *ROracle* and *DBI*

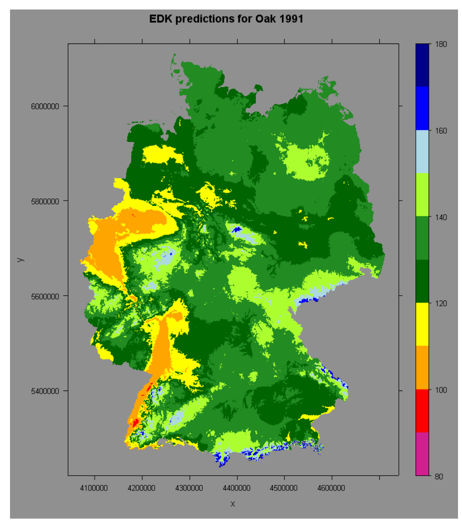
Computed budburst date based on NDVI values

NDVI, 1989-1997, Atmospheric filtering and corrections (Koslowsky et al. 2001, Koslowsky et al. 2003)
Dynamic filtering BISE (Viovy et al. 1992)

Phenological observation stations in Germany



Spatial interpolation of ground observations

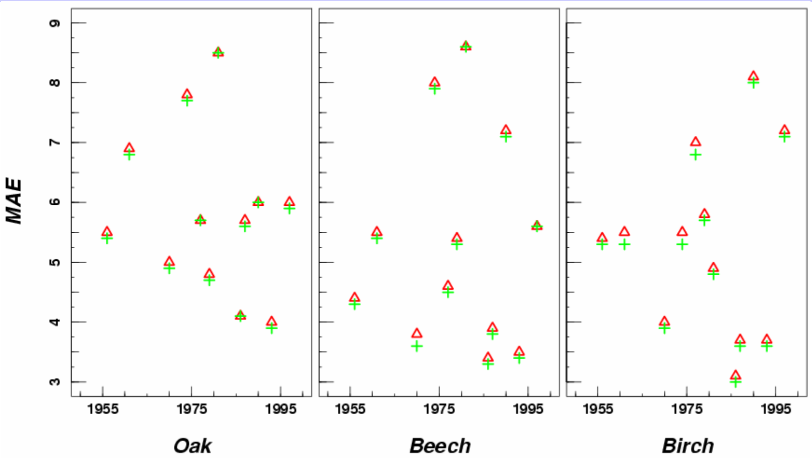


- External Drift Kriging (EDK), thereby incorporating elevation as secondary information
- Detrended Kriging based on a Global elevation gradient:

$$g_{ha} = \frac{\sum_{i=1}^{n} \left(\frac{\Delta d_{obs}}{\Delta h}\right)_i * w_i}{\sum_{i=1}^{n} w_i}$$

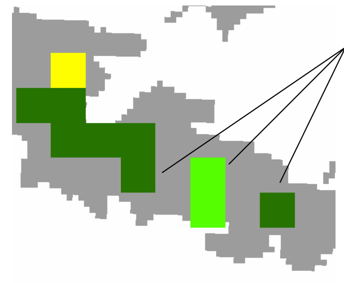
- Used package for geostatistical purposes: *gstat*

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

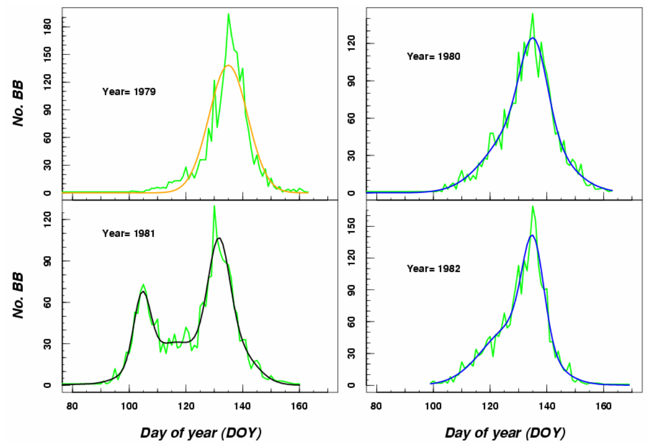
$f_1(x), \ldots, f_m(x)$ the component densities

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

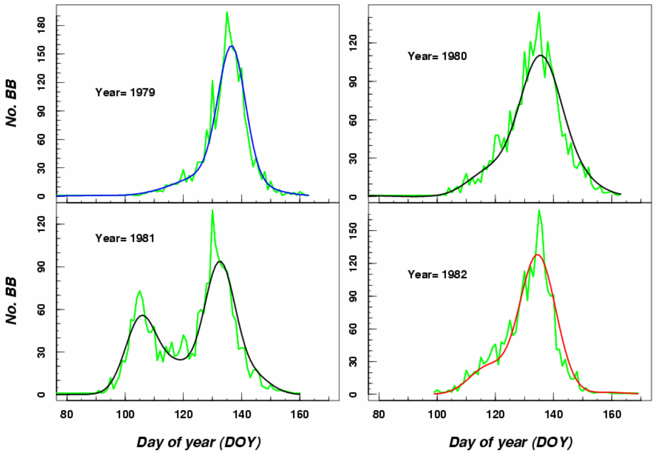


Frequency distributions of observations of Oak (*green* line) modelled using between 1-4 Mixture distributions

orange = 1 component(s)
blue = 2 ""
red = 3 ""
black = 4 ""

	μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4	ω_1	ω_2	ω_3	ω_4
1979	134.95				6.78				1			
1980	132.09	135.33			11.86	4.70			0.67	0.33		
1981	104.54	116.31	131.55	137.51	2.94	11.34	3.83	6.78	0.15	0.37	0.32	0.16
1982	128.15	135.14			10.68	3.78			0.62	0.38		

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

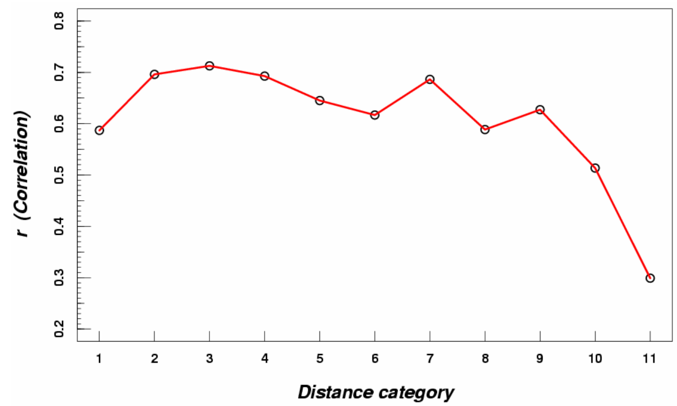
- Time series' correlation of station pairs is assigned to distance categories

1	2	3	4	5	6	7	8	9	10	11	
0-25	25-50	50-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800	800-900	[km]

- 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*

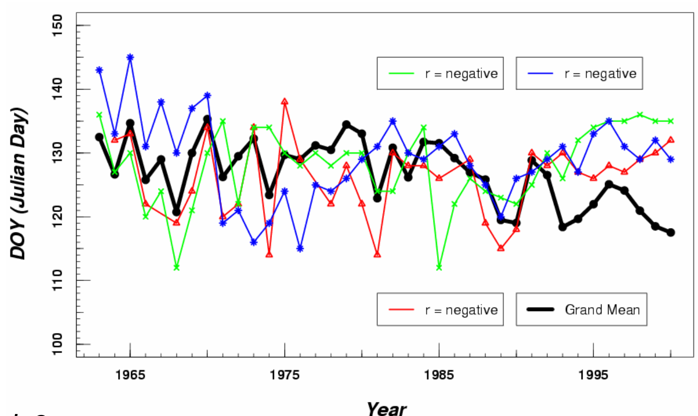


after Koenig et al. (1998)

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



Origin of reversed trends?

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

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 spring time.

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