Smoothed Prediction of the Onset of Tree Stem Radius Increase Based on Temperature Patterns

Mikko Korpela[1] <Mikko.V.Korpela@tkk.fi> Harri Mäkinen[2]
Mika Sulkava[1] Pekka Nöjd[2] Jaakko Hollmén[1]

[1]Helsinki Institute for Information Technology
Helsinki University of Technology, Finland
Department of Information and Computer Science

[2]Finnish Forest Research Institute

TIES 2008 – Kelowna, Canada

Outline

Introduction

- Changes in climate all over the world
- Forest industry is important in Finland
- We want to model relationship between two processes:
  - Environmental factors
  - Growth of trees
- The problem setting (don't try to solve everything):
  - Predict yearly onset date of radial stem increase
  - Using only temperature information

Traditional Temperature Sums

- Keep a record of daily temperatures
- Take a cumulative sum
- Only count the part exceeding a threshold, e.g. $+5\,^{\circ}\mathrm{C}$
- Can be used as an explanatory factor for growth
- Method does not use data very well
    - Temperature time series summarized with one number

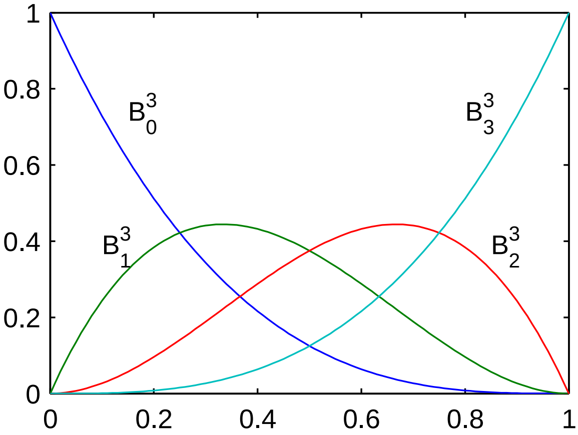
Extracting More Information from Weather Data

Temperature Features with Bernstein Polynomials

- We consider a weather history of $N = 80$ days
- Several temperature features are computed
- Each feature represents a different part of history
- Temperatures are weighted with Bernstein basis polynomials
- Plain temperatures are used ($+5\,^{\circ}\mathrm{C}$ threshold not used)
- All features considered, the whole history weighted equally
- If features are dropped, some locations (e.g. recent observations) will get more relative weight

Bernstein polynomials

Example: Polynomials of Degree 3



Extracting More Information from Weather Data

The Technical Part

Bernstein Basis Polynomials of Degree d

$$B_i^d(x) = \binom{d}{i} x^i (1-x)^{d-i} \,, \quad i = 0, \ldots, d \,.$$

Temperature Features

$$s_m(i) = \sum_{j=1}^{N} B_m^d\left(\frac{j-1}{N-1}\right) T_{i-j} \,, \quad m = 0, \ldots, d \,.$$

- Number of features is adjusted by changing the degree d
- N is length of history window
- T_i is temperature on day i

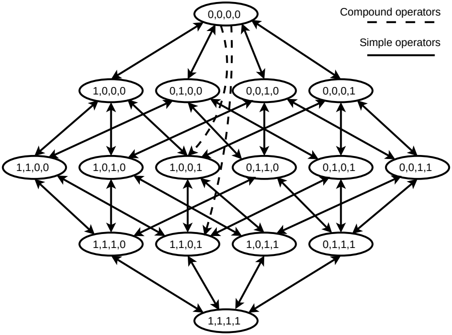
Selecting Temperature Features

- The Bernstein-weighted temperature features are correlated
- Possible benefits by using a subset of features:
  - Model easier to understand
  - Better performance in prediction task
- We use the Best First Search (BFS) feature selection algorithm
- Non-exhaustive state-space search
- Each state represents a set of features
- Search is guided by the performance of the prediction machine when using each set of features

Selecting Temperature Features (2)

Structure of state space in BFS

New states are evaluated by expanding the best known state (add or remove one feature, optional compound operators).



The Prediction Machine

Input

- Temporally localized temperature features for the current date

Output

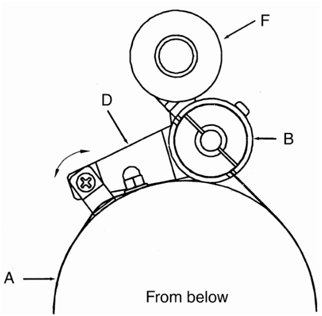
- Predicted time until onset of radial growth (days)

The Model

- Linear combination of parametric and non-parametric models: linear regression and k-NN
- Prediction sequence is smoothed by combining old prediction and novelty

Dendrometer Data

- Stainless-steel band for measuring stem circumference
- 57 year × tree combinations from 2001–2005 (Southern Finland)
- High-resolution measurements stored as 1-h averages
- Further conversion to daily magnitude of radial change
- Onset date of radial change determined visually

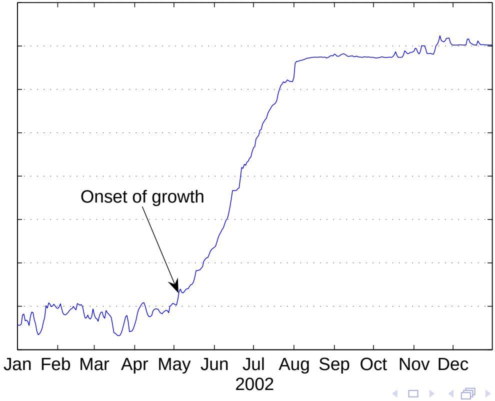


A Stainless-steel band
B Rotating potentiometer
D Adjustable foot
F Spring

Example: Dendrometer Data

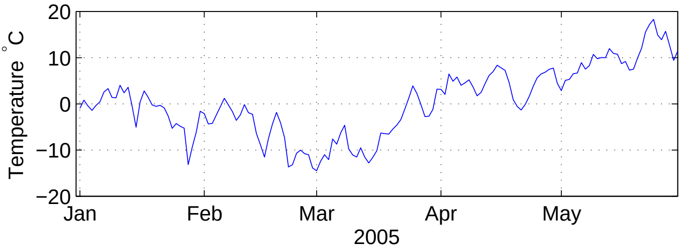
Visually Determined Onset Date Marked

Automatic detection of onset date possible with CUSUM chart (Sulkava et al. TIES 2007)



Temperature Data

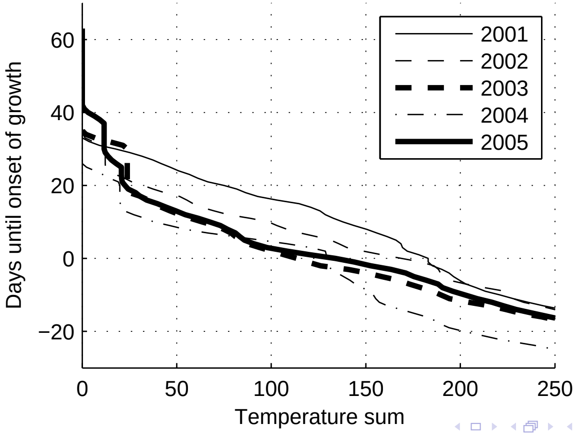
- Finnish Meteorological Institute
- Measurement site about 5 km from growth sites
- 3-hour measurement intervals
- Measurements averaged to daily values for our purposes



Motivation for the Temperature Features

Temperature Sum Alone is a Poor Predictor

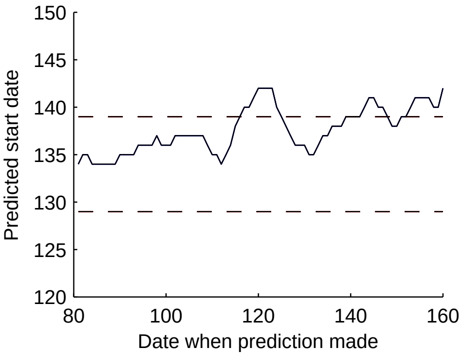
Variation between years is large (see picture)



Example Output of the Prediction Machine

Prediction vs. Date When Prediction Made

Staying between horizontal dashed lines is desired



Prediction Accuracy

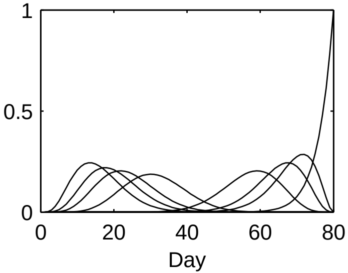
- 4-fold cross-validation tests
- Errors are in days
- Only k-NN or both linear regression and k-NN
- Results superior to k-NN on temperature sum (bottom row)
- Large year-to-year variation in accuracy (column “Test”)

	#Features			RMSE		
Model	Chosen	Total	k	Valid.	Test	(std)
lin + k-NN	16	40	5	5.2	5.5	(1.2)
k-NN	16	40	35	5.3	5.7	(1.4)
lin + k-NN	8	20	5	5.4	6.9	(1.6)
k-NN	9	20	25	5.4	4.1	(1.0)
lin + k-NN	40	40	90	5.7	6.4	(1.6)
lin + k-NN	20	20	105	5.8	6.5	(1.5)
k-NN on temperature sum			50	8.5	3.2	(0.8)

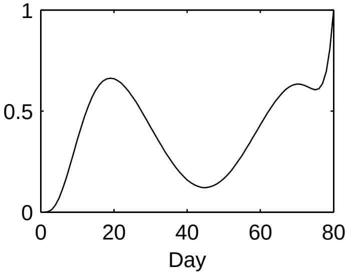
Selected Temperature Features

8 out of 20 features

Features selected with BFS. X-axis: large number means distant past.



Individual weighting functions



Sum of weights

Summary

- New kind of temporally localized temperature features for predicting onset of tree growth
- Combination of a parametric and a non-parametric regression method
- Improved accuracy compared to traditional degree-days

- Outlook
  - Look for a possible trend in the past onset dates
  - Apply methodology to other problems