Linear Predictive Coding and Cepstrum coefficients for mining time variant information from software repositories

G. Antoniol, F. Rollo and G. Venturi
RCOST – University of Sannio - Italy
LPC Idea

- Model a time series with a polynomial approximation
  - LPC Cepstrum
  - smooth the spectrum
- Define the distance between two time series as the distance between their polynomial approximations
- Use distance to cluster time series with identical or similar evolutions.
LPC and Linux Kernel

- 211 Linux releases about 1700 files
- Study the influence of the number of coefficients
- Study the influence of distance thresholds
- Mine files with similar evolution:
  - Create groups of files with the same or very similar size evolution