Program of time series workshop

	Monday at 2pm		Tuesday at 1 pm	
	Main Room		**Main Room**	
15 min	Prof. Bart De Moor: Introduction		Prof. Bart De Moor: Introduction	
10 min	Edward De Brouwer		Lola Botman	
10 min	Jonas Soenen		Nick Seeuws	
	Room B1 moderator Philippe Dreesen	**Room B3** moderator Katrien De Cock	**Room B1** moderator Philippe Dreesen	**Room B3** moderator Katrien De Cock
20 min	discussion 1 Edward De Brouwer	discussion 1 Jonas Soenen	discussion 1 Lola Botman	discussion 1 Nick Seeuws
20 min	discussion 2 Jonas Soenen	discussion 2 Edward De Brouwer	discussion 2 Nick Seeuws	discussion 2 Lola Botman

AI FLANDERS

AI for Time Series

Prof. Bart De Moor
ESAT-STADIUS, KU Leuven

AI Flanders research days April 2021

AI VLAANDEREN
BOUWEN AAN JE DIGITALE TOEKOMST

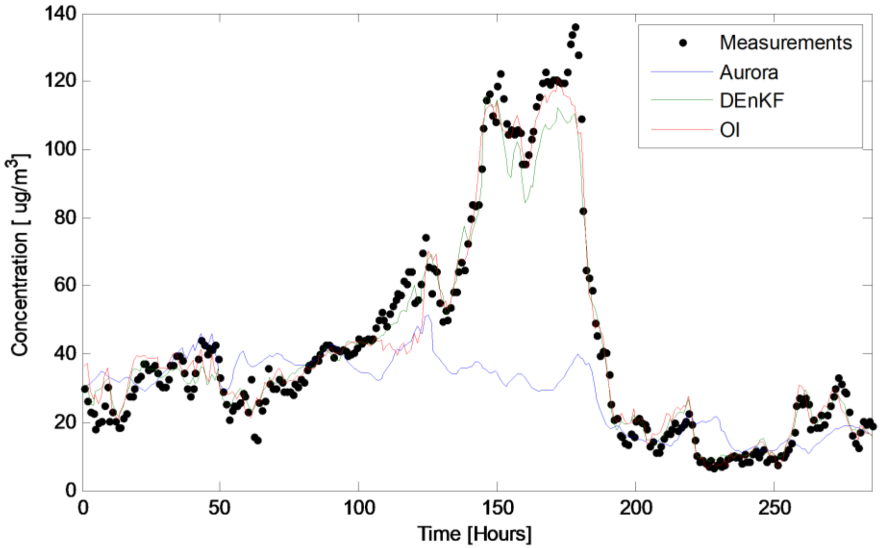
AI FLANDERS
BUILDING OUR DIGITAL FUTURE

Outline

- What is a time series?
- Motivation for time series analysis
- Time series analysis
- Historical notes
- Dynamics
- Distance measures for time series
- Extensions
- The modeling cycle
- Results of e-mail survey

What is a time series?

- time series = collection of observations made sequentially in time
- order of observations is important
- causality: the past influences the future



measurements of particulate matter and estimates existing model: Aurora, improved models: DEnKF (ensemble Kalman filter) and OI (Optimal Interpolation)

Motivation for time series analysis

diagnose past behavior, predict future behavior

- forecasting
- simulation
- clustering and classification
- anomaly detection
- use of model in control loop
- segmentation

Time series analysis

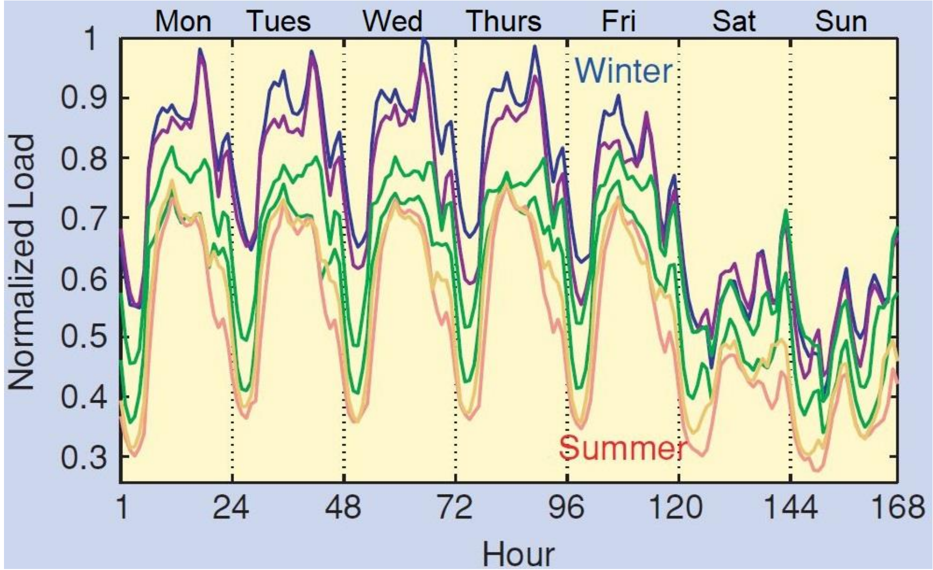
constructing models from observed time series

- time domain vs frequency domain
- parametric vs non-parametric
- linear vs non-linear
- univariate vs multivariate

Time series analysis

example: electric load forecasting

- data: hourly load values from substations in the Belgian grid
- seasonality: day, week, year
- inputs of model:
  - seasonal patterns
  - temperature



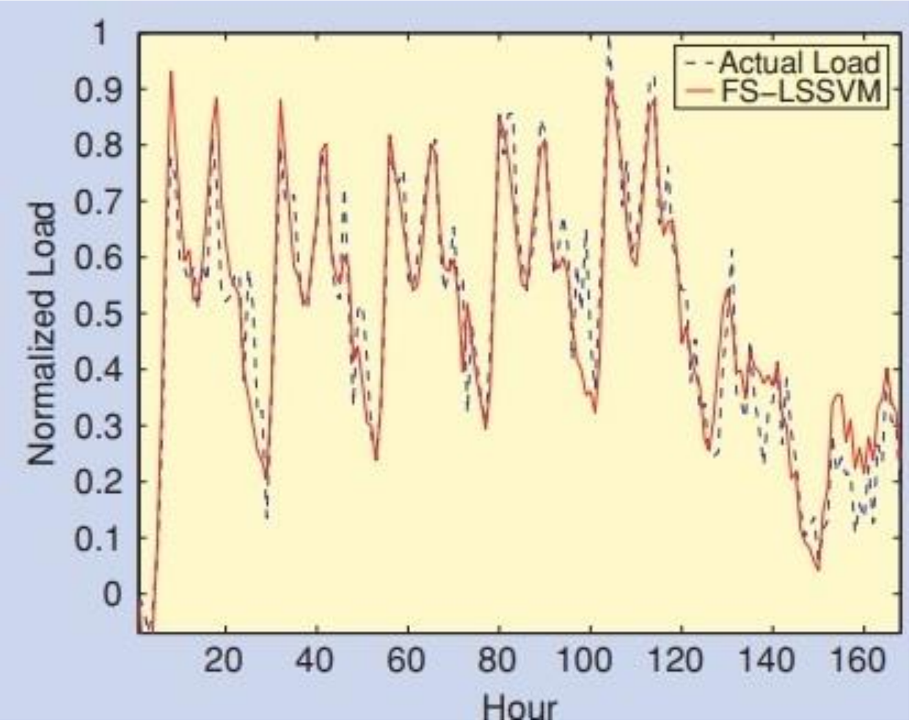
Weekly profile over the year. The load in winter (blue) is different from the load in summer (red), both being different from profiles during spring or autumn (green).

Source: M. Espinoza, J.A.K. Suykens, R. Belmans, B. De Moor, Electric Load Forecasting - Using kernel based modeling for nonlinear system identification, IEEE Control Systems Magazine, Special Issue on Applications of System Identification, vol. 27, no. 5, Oct. 2007, pp. 43-57.

Time series analysis

- model: NARX = nonlinear autoregressive with exogenous inputs
- nonlinearity modeled by FS-LSSVM = fixed size least squares support vector machine

forecasting result



measured load (blue dashed line) and 24h-ahead forecasting (red line)

Time series analysis: increasingly important

- massive production of time series data
  - digitization of health care
  - internet of things
  - smart cities
  - smart electricity meters
  - process industry
  - telecommunications
  - music
  - weather and climate
  - sports
  - finance
- continuous monitoring and data collection
- increased computing power

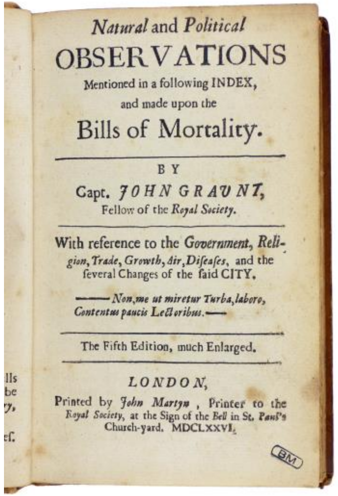
Historical notes

contributions from variety of application areas

- astronomy: orbit estimation from observations
- demography: actuarial tables
- medicine: ECG, EEG
- weather forecasting
- forecasting economic growth
- financial markets

Johannes Kepler
(1571-1630)

J. C. F. Gauss
(1777-1855)



Natural and *Political*
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

BY
Capt. *JOHN GRAUNT*,
Fellow of the *Royal Society*.

With reference to the *Government*, *Religion*, *Trade*, *Growth*, *Air*, *Difeafes*, and the feveral Changes of the faid CITY.

——*Non, me ut miretur Turba, laboro,*
Contentus paucis Lectoribus.——

The Fifth Edition, much Enlarged.

LONDON,
Printed by *John Martyn*, Printer to the *Royal Society*, at the Sign of the *Bell* in St. *Paul's* Church-yard. MDCLXXVI.

John Graunt
(1623-1687)

Historical notes

Development of theory: different scientific communities

1. statistics
2. econometrics
3. systems and control

Historical notes

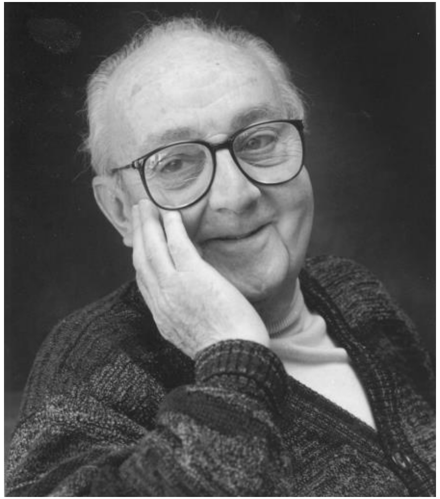
1. statistics

   time series = stochastic process

   - Yule (1871-1951) application of AR model to sunspot data
   - Fisher (1890-1962)
   - Box (1919-2013) "All models are wrong but some are useful"
   - Akaike (1927-2009)

2. econometrics

   - Tinbergen (1903-1994)
   - Hannan (1921-1994)
   - Granger (1934-2009)
   - Hamilton (°1954)



George Box

Historical notes

- systems and control
  time series = output of dynamical system
  modeling of time series = system identification
  - Kalman (1930-2016)
  - Willems (1939-2013)
  - Ljung (°1946)

Dynamics

dynamical system:
current output depends not only on current external stimuli, but also on their earlier values

- static: algebraic equation
- dynamical:
  - differential/difference equation
  - state space equations

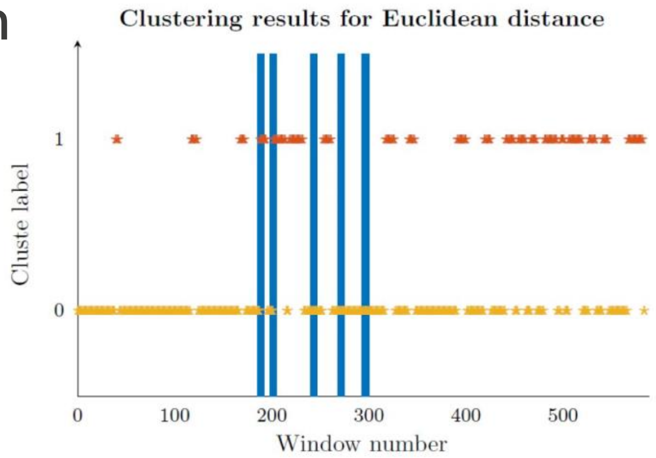
Distance measures for time series

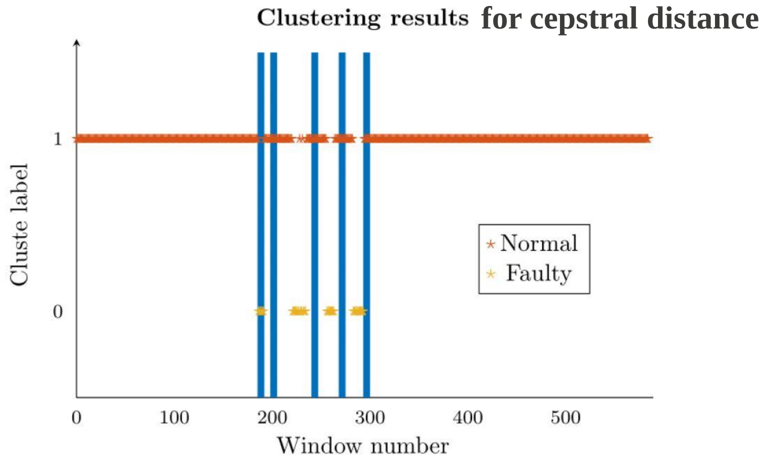
- useful for: clustering, classification, anomaly detection
- how similar are the systems generating the time series?
- important: include dynamics
  - autocorrelation
  - spectrum
  - cepstrum
  - dynamical model parameters

Distance measures for time series

example: clustering for anomaly detection

- data:
  - wind turbine
  - input: wind speed
  - output: temperature of stator
  - sliding windows: length = 1 week
- clustering algorithm: k-means
- distance measure: Euclidean vs cepstral
- cluster 0 = yellow
  cluster 1 = red
- blue: error logs about recent faults
- cepstral: only first fault is missed
- Euclidean: no distinction normal/faulty behavior

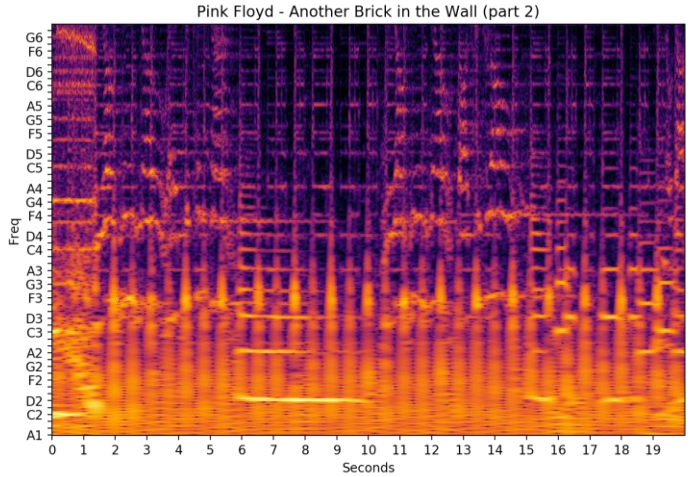


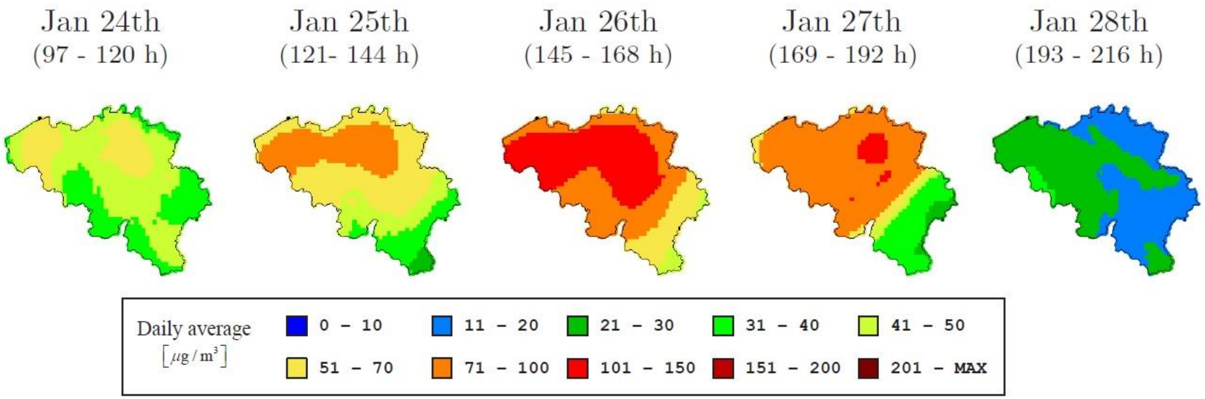


Extensions of time series

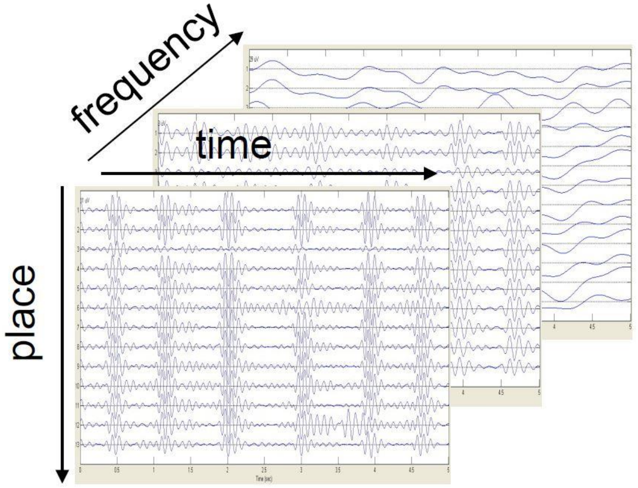
mutidimensional time series

- air quality: place and time
- music: frequency and time
- EEG signals: place, time and frequency





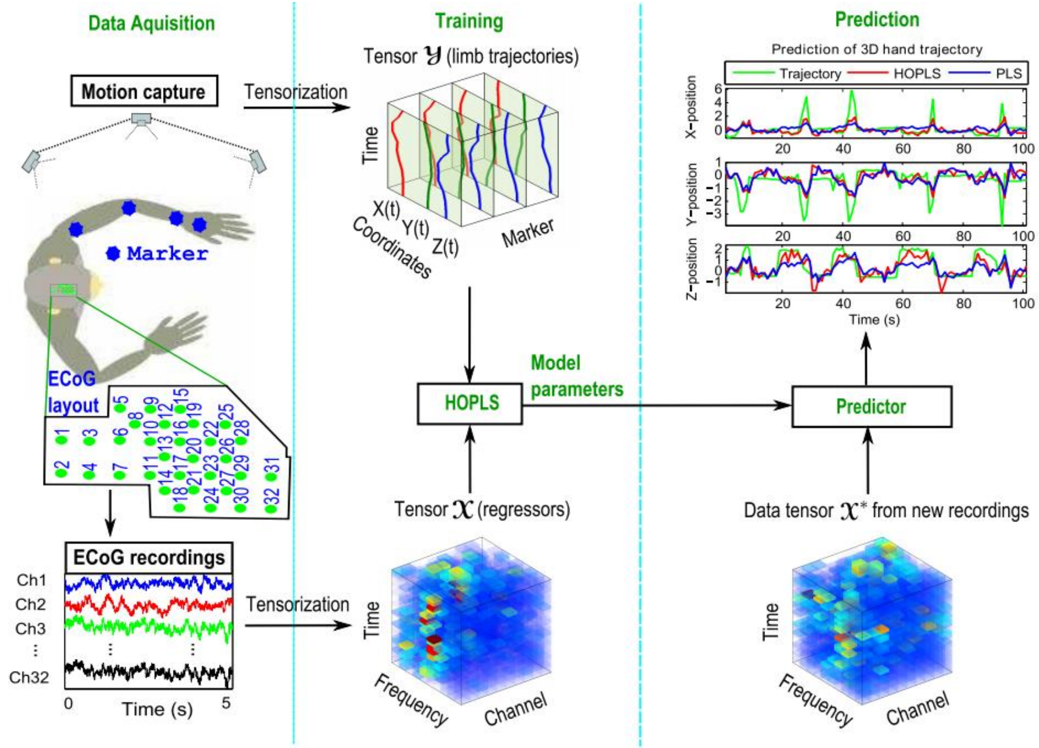
Daily average PM_{10} concentration for January 24th -January 28th, 2010. Optimal Interpolation method.



Source: Sabine Van Huffel, The power of tensor algebra in medical diagnosis

Extensions of time series

tensor based data analysis

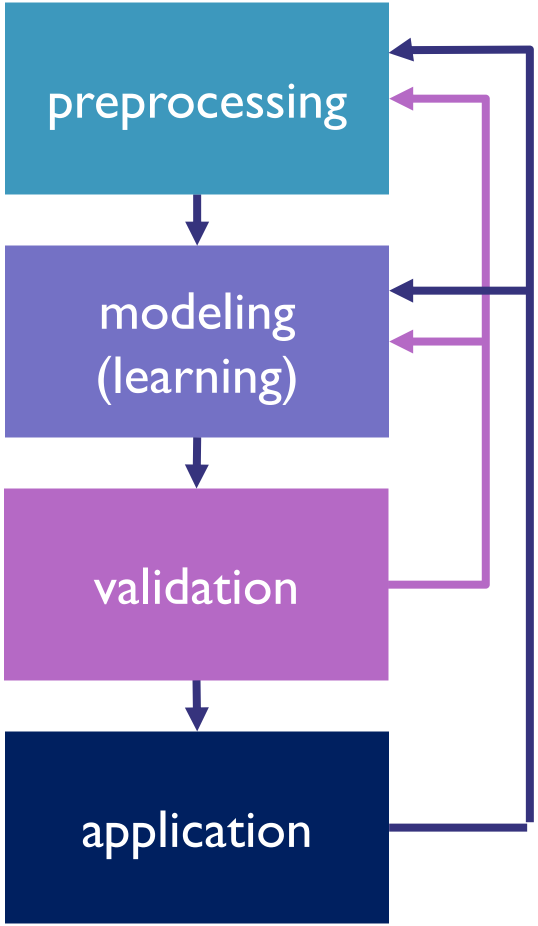


Prediction of arm movement from brain electrical responses.
Left: Experiment setup.
Middle: Construction of the data and response tensors and training.
Right: The new data tensor (bottom) and the predicted 3D arm movement trajectories (X, Y, Z coordinates) obtained by tensor-based HOPLS and standard matrix-based PLS (top)

Source: A. Cichocki, D. P. Mandic, A. H. Phan, C. F. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, (2015), Tensor Decompositions for Signal Processing Applications. From two-way to multiway component analysis IEEE Signal processing Magazine, 32 (2), 145-163.

Modeling cycle

Aspects of time series analysis that will be addressed in the discussion sessions



Modeling cycle: user's choices

1. Preprocessing

- remove trend, seasonality, ...
- resample, subsample, oversample
- remove outliers
- deal with missing data
- filter data
- extract features

Modeling cycle: user's choices

2. Modeling

- non-parametric or parametric
- linear or non-linear
- model structure
- number of parameters
- identification/optimization criterion

Modeling cycle: user's choices

3. Validation

- physical parameters feasible?
- check if 'true system' is in model class: correlation tests
- look at variance of parameters
- model reduction possible?
- k-step ahead prediction
- simulation
- compare measured signal with predicted/simulated signal using different data set

Modeling cycle: user's choices

4. Applications

- medical: many challenges
  - high quality data: costly to acquire
  - often low sample sizes
  - large proportion of missing values
  - often imbalanced longitudinal data (irregularly sampled)
- finance
  - fraud detection
  - portfolio management

Modeling cycle

- smart cities:
  - predictive maintenance
  - flood prediction
  - traffic management
  - electricity demand

Results of e-mail survey

Topics of interest to workshop's participants

1. Preprocessing
   feature extraction, auto-encoder, segmentation, data visualization, choosing sampling interval, resampling, subsampling, automation of preprocessing, peak detection, irregularly sampled TS, transformation to symbolic strings, complex-valued signals, detection of quality issues, information compression
2. Modeling - Learning
   clustering, classification, anomaly detection, model choice (explainable or black-box), finding optimal hyperparameters, make models more explainable, detection of periodicity, interpretability of methods, random forests, statistical models, quantify uncertainty for extrapolation, LSTMs, representation learning, neural networks, hierarchical representation, model reduction

Results of survey

3. Validation

   expert feedback to improve methods

4. Application

   - electricity: forecasting of electricity consumption, clustering for electricity consumption, forecasting peaks in electricity consumption
   - medical: sleep apnea, multiple sclerosis, prediction of length of stay of hospital patients, detection of epileptic seizures
   - understand and interpret brain signals (EEG, fNIRS)
   - mechanical machinery (bearings, gearboxes) condition monitoring
   - audio and music
   - agent behavior in grid world

5. Software and tools

   to be added in discussion session

Program of time series workshop

	Monday at 2pm		Tuesday at 1 pm	
	Main Room		**Main Room**	
15 min	Prof. Bart De Moor: Introduction		Prof. Bart De Moor: Introduction	
10 min	Edward De Brouwer		Lola Botman	
10 min	Jonas Soenen		Nick Seeuws	
	Room B1 moderator Philippe Dreesen	**Room B3** moderator Katrien De Cock	**Room B1** moderator Philippe Dreesen	**Room B3** moderator Katrien De Cock
20 min	discussion 1 Edward De Brouwer	discussion 1 Jonas Soenen	discussion 1 Lola Botman	discussion 1 Nick Seeuws
20 min	discussion 2 Jonas Soenen	discussion 2 Edward De Brouwer	discussion 2 Nick Seeuws	discussion 2 Lola Botman