Adaptive learning algorithms for data streams

Dr Patrick McSharry

patrick@mcsharry.net

Centre for Catastrophe Risk Financing
Smith School of Enterprise and the Environment
University of Oxford
www.smithschool.ox.ac.uk/crf

DPhil students:
Siddharth Arora and Shilan Mistry

15th February 2010
Bristol Algorithms Days 2010 Feasibility Workshop

University of Bristol / Heilbronn Institute

Outline

- Quantitative decision-making
- Simplicity and complexity
- Modelling and uncertainty
- Ensemble modelling
- Adaptive models
- Model combinations

Human experts versus machines

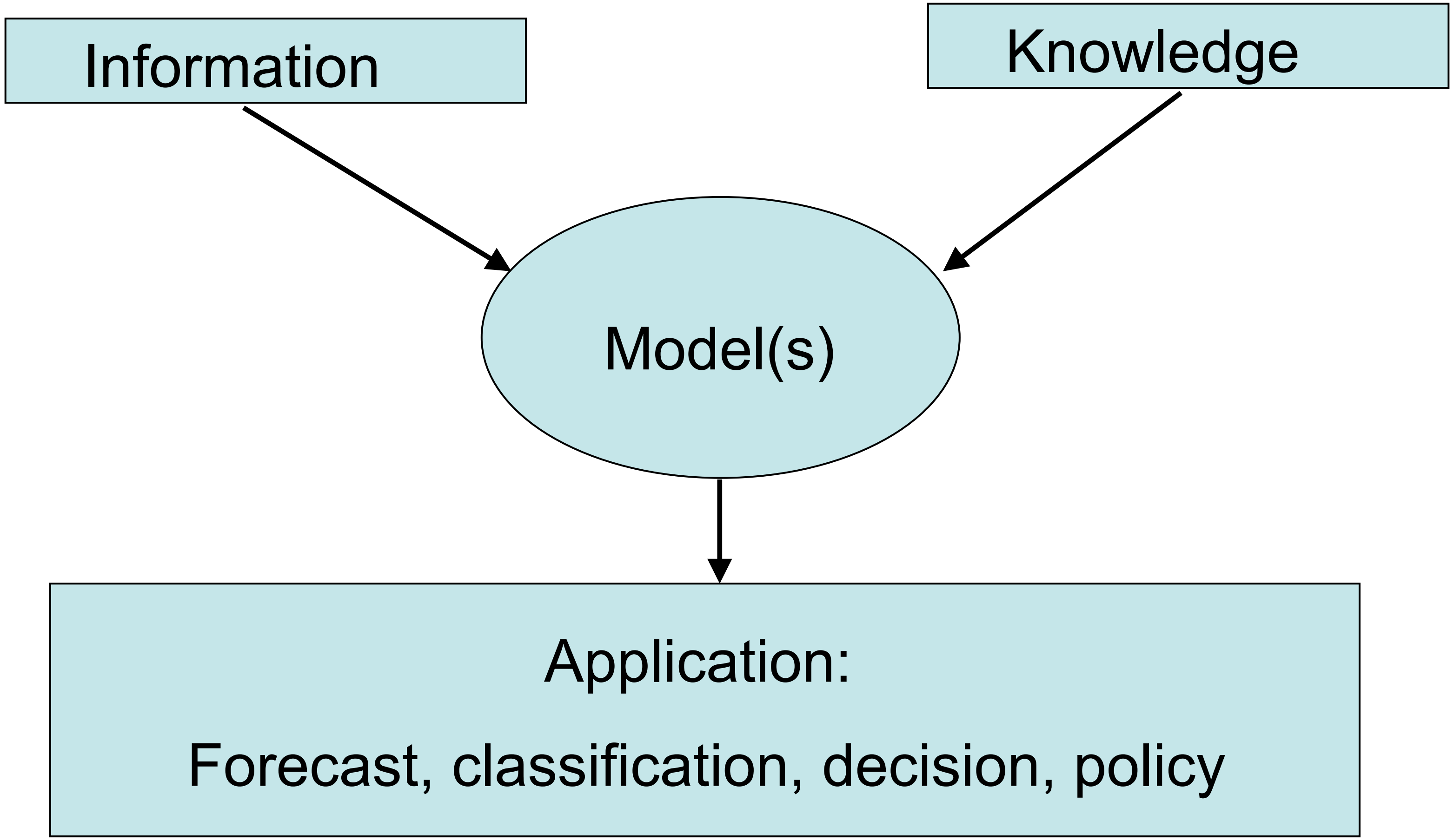
- Traditionally important decisions have been made by human experts
- Prediction analytics due to availability of data, computational resources and machine learning/statistical techniques
- Accounting for uncertainty is key for quantifying the confidence underlying the decision-making process

Quantitative decision-making

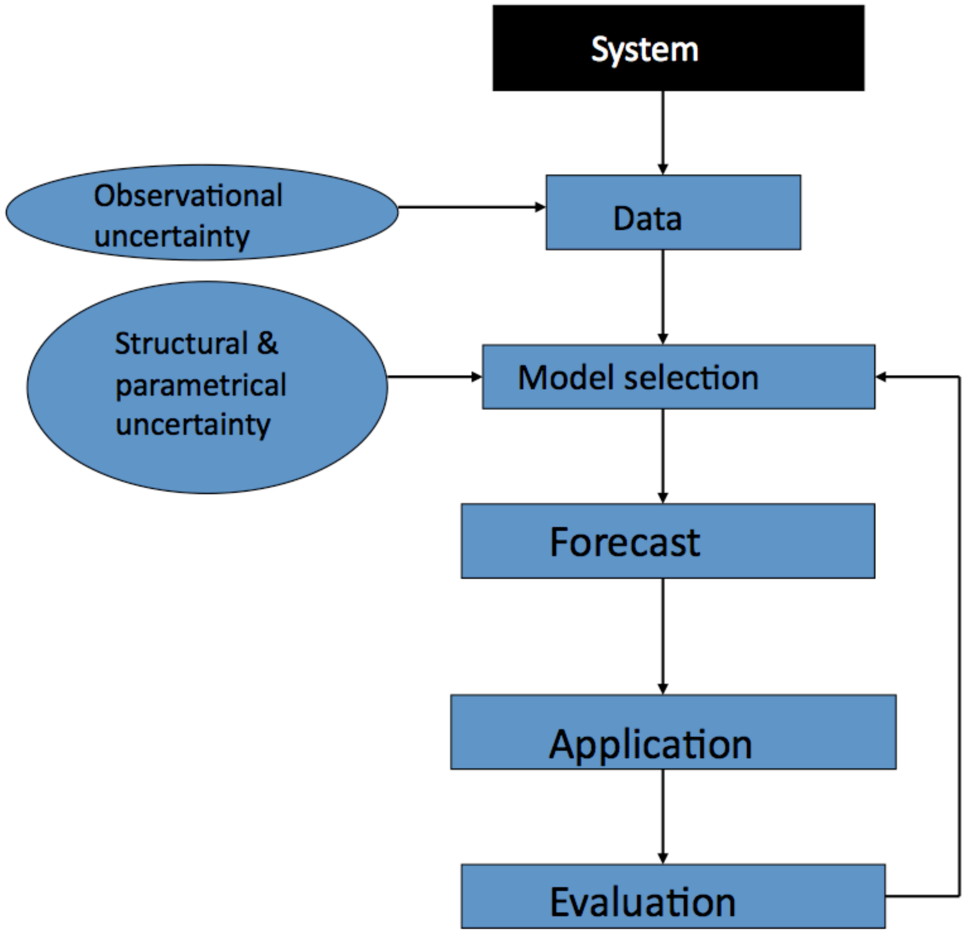
- Challenges:
  - Observational uncertainty, errors
  - Nonstationarity, external shocks

- Advantages:
  - Consistency and robustness
  - Data streaming and adaptive algorithms

Model construction



Modelling & Uncertainty

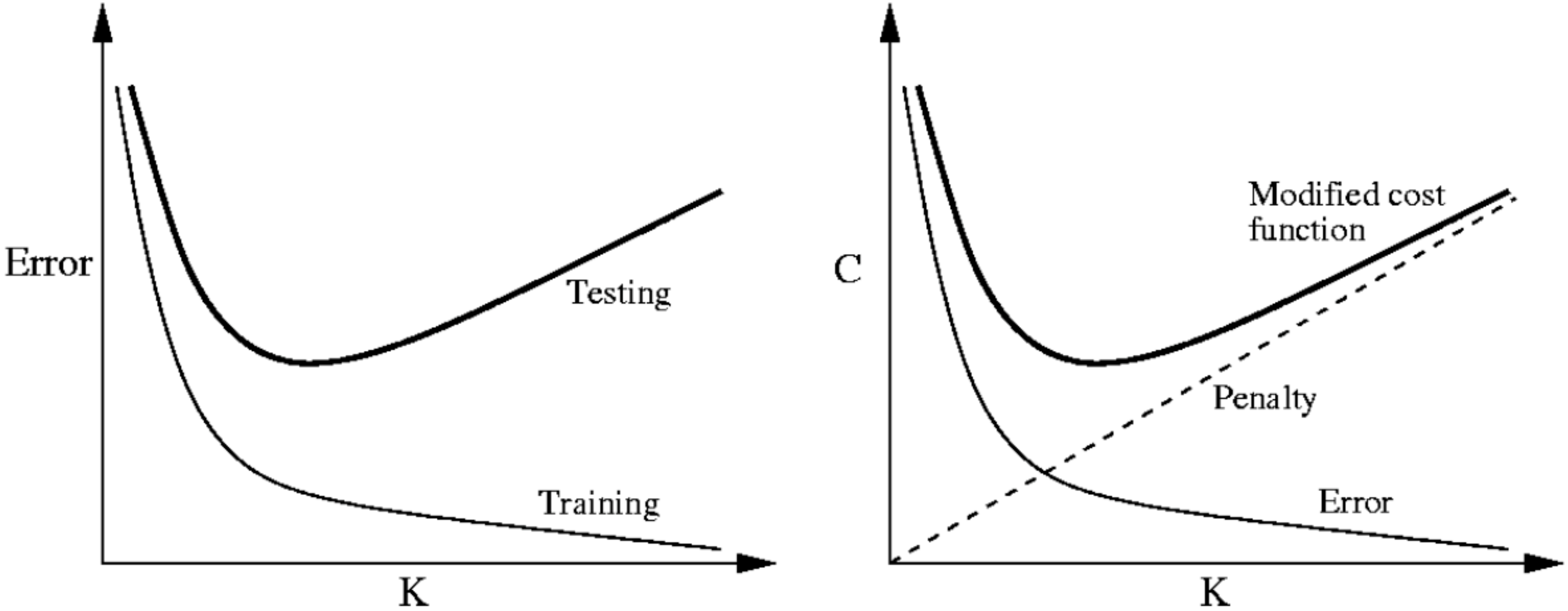


Occam's Razor

William of Occam studied theology at the University of Oxford from 1309 to 1321, but never completed his master's degree

- Occam's razor is a principle attributed to the 14th-century English logician and Franciscan friar William of Ockham
- The principle states that a theory should rely on as few assumptions as possible, eliminating those that make no difference to the observable predictions of the theory
- Given multiple competing theories that are equally plausible, the principle of Occam's Razor suggests selecting the theory that relies on the fewest assumptions
- "Everything should be made as simple as possible, but not simpler" - Einstein

Parsimony and model selection



An obvious sign of model over-fitting is one that performs better on training (in-sample) data than testing (out-of-sample) data. Implemented via AIC and BIC.

Source: McSharry (2005), The danger of wishing for chaos

Risky simplifications

Carl Friedrich Gauss
1777 - 1855

Louis Bachelier
1870 - 1946

- Gaussian distributions
- Random walk and independence
- Analytical elegance and convenience
- Value at risk – static description of risk based on history
- Regulators extended VaR: 95% to 99.5% and one day to one year
- "Events that models only predicted would happen once in 10,000 years happened every day for three days" - an employee of Lehman Brothers on August 11, 2007
- Insurance extreme events: Hurricane Katrina & Andrew

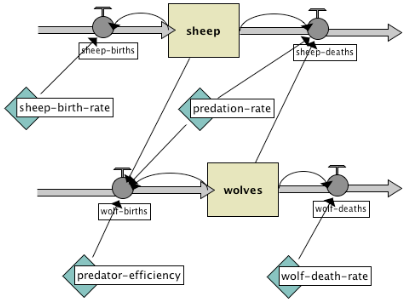
Nonlinear dynamics

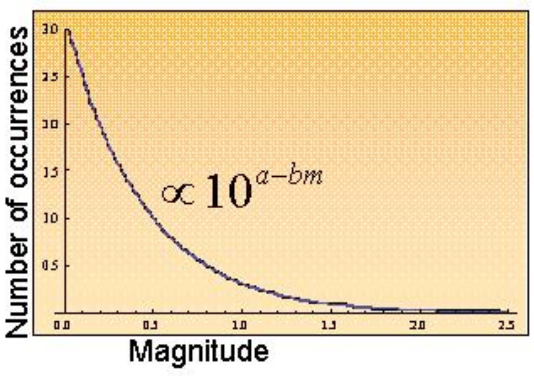
- Nonlinearity may result from regime switching and feedback loops
- Nonlinear systems may experience chaos or sensitivity to initial conditions – this can be accounted for using ensemble predictions
- Nonlinearity implies non-normal distributions and the need for quantile forecasts for risk analysis

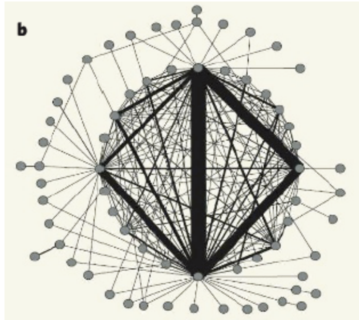
Forecast failure

- Despite the improvement in data quality and computer processing, traditional forecasting models have proved disappointing in applications to complex systems such as the weather, genetics, finance and economics
- These systems are characterised by properties such as nonlinearity, connected networks, multiple scales, and emergent properties
- Isaac Newton noted in 1721, after he lost most of his fortune in the collapse of the South Sea bubble: "I can calculate the motions of heavenly bodies, but not the madness of people."

Systems modelling







- Systems modelling is an approach that explicitly takes account of these properties
- It takes advantage of mathematical techniques and concepts that have been developed for complex systems, such as agent-based models, network analysis, and system dynamics

Orrell & McSharry (2009), International Journal of Forecasting 25(4): 734-743.

Systems forecasting

- The systems approach does not aim to provide a single "theory of everything."
- Models are seen as imperfect, fuzzy patches that can be adapted for particular situations – each revealing a different, but incomplete aspect of the entire system.
- The appropriate modeling technique for a complex system often depends on the scale (granularity) being addressed and the question being asked.
- Models of complex systems need not be complicated; indeed parsimonious models often provide superior forecasts if they take into account the relevant properties.

Orrell & McSharry (2009), International Journal of Applied Forecasting (Foresight)14: 24-30.

Electricity demand forecasts

- Demand is important for estimating capacity required to avoid black-outs
- Control of generation and distribution
- Ability to make informed decisions
- Reduce risks and minimise costs
- Implementation depends on the forecast horizon:
  - Short-term: ensuring system stability
  - Medium-term: maintenance scheduling
  - Long-term: capital planning
- Required for studying electricity prices
- Design of efficient electricity markets

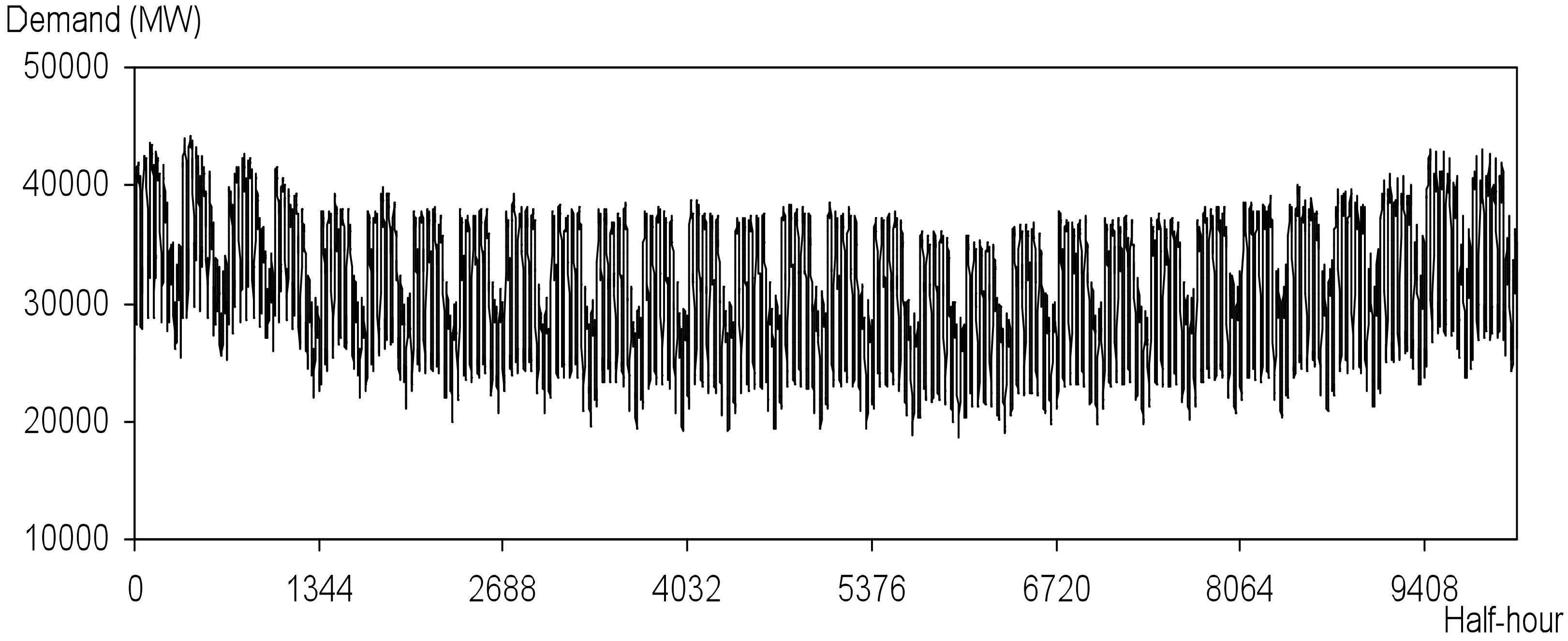
Influences on energy demand

- Deterministic seasonality:
  - daily and weekly seasonality (low demand at weekends)
  - Annual seasonality (low demand during the summer)
- Calendar effects:
  - Summer holidays, Christmas holidays
  - Bank holidays
  - Bridge days, elections, strikes, …
- Weather effects:
  - Temperature
  - Wind speed
  - Luminosity

McSharry, Bouman & Bloemhof (2005), IEEE Trans Power Systems 20(2): 1166-1172.

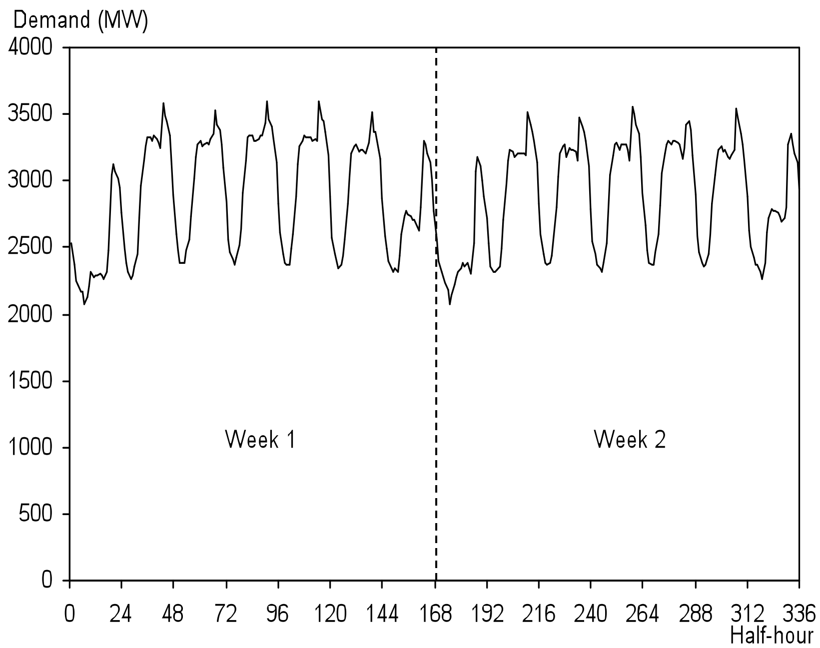
England and Wales demand

Within-day cycle (48 periods) and a within-week cycle (336 periods).

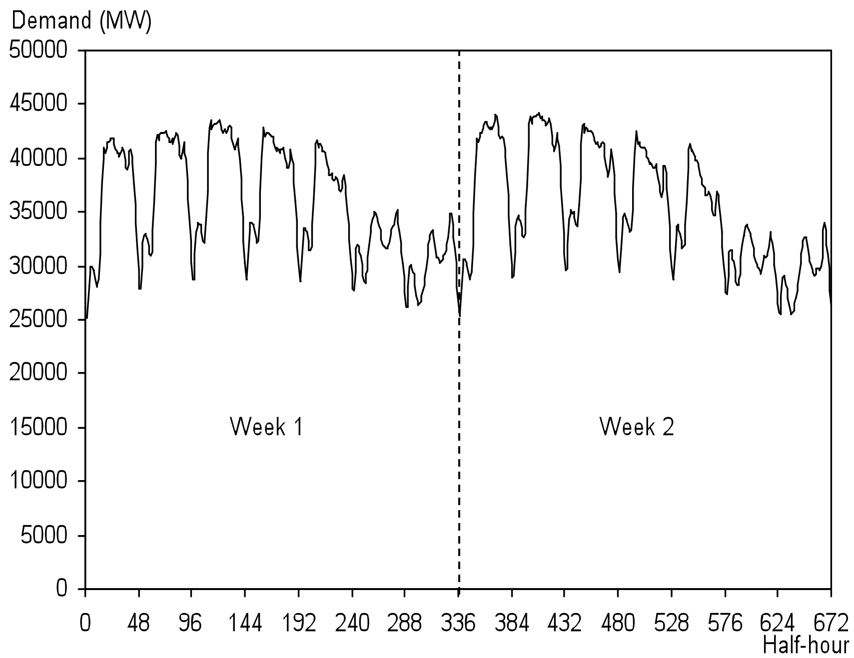


Monday 27 March 2000 to Sunday 22 October 2000

Two weeks of demand



Rio



England
& Wales

Moving averages

- A moving average may be viewed as a convolution or a low-pass filter.
- A simple moving average of the last n observations is given by:

$$s_t = \frac{1}{n}\sum_{i=1}^{n} x_{t-i+1}$$

- By smoothing the time series, it can remove the effects of seasonality.
- Note that it gives equal weight to both new and old observations

Exponentially weighted moving averages

- EWMA employs exponentially decreasing weights to discount the influence of old observations:

$$s_t = \alpha x_t + (1-\alpha)s_{t-1}$$

- The smoothing factor, $0<\alpha<1$, determines the rate of decay of old information
- The EWMA may also be expressed via the number of time periods n using the smoothing factor, $\alpha=1/n$, *giving:*

$$s_t = \left(\frac{1}{n}\right)x_t + \left(\frac{n-1}{n}\right)s_{t-1}$$

Double Seasonal Exponential Smoothing

- Extension of Holt-Winters:

$$\hat{y}_t(k) = (S_t + k\,T_t)\,D_{t-s_1+k}\,W_{t-s_2+k}$$

  - Level: $S_t = \alpha\,(X_t/(D_{t-S_1}W_{t-S_2})) + (1-\alpha)(S_{t-1} + T_{t-1})$
  - Trend: $T_t = \gamma\,(S_t - S_{t-1}) + (1-\gamma)\,T_{t-1}$
  - Seasonality 1: $D_t = \delta\,(X_t/(S_t W_{t-S_2})) + (1-\delta)\,D_{t-S_1}$
  - Seasonality 2: $W_t = \omega\,(X_t/(S_t D_{t-S_1})) + (1-\omega)W_{t-S_2}$

Double Seasonal Exponential Smoothing

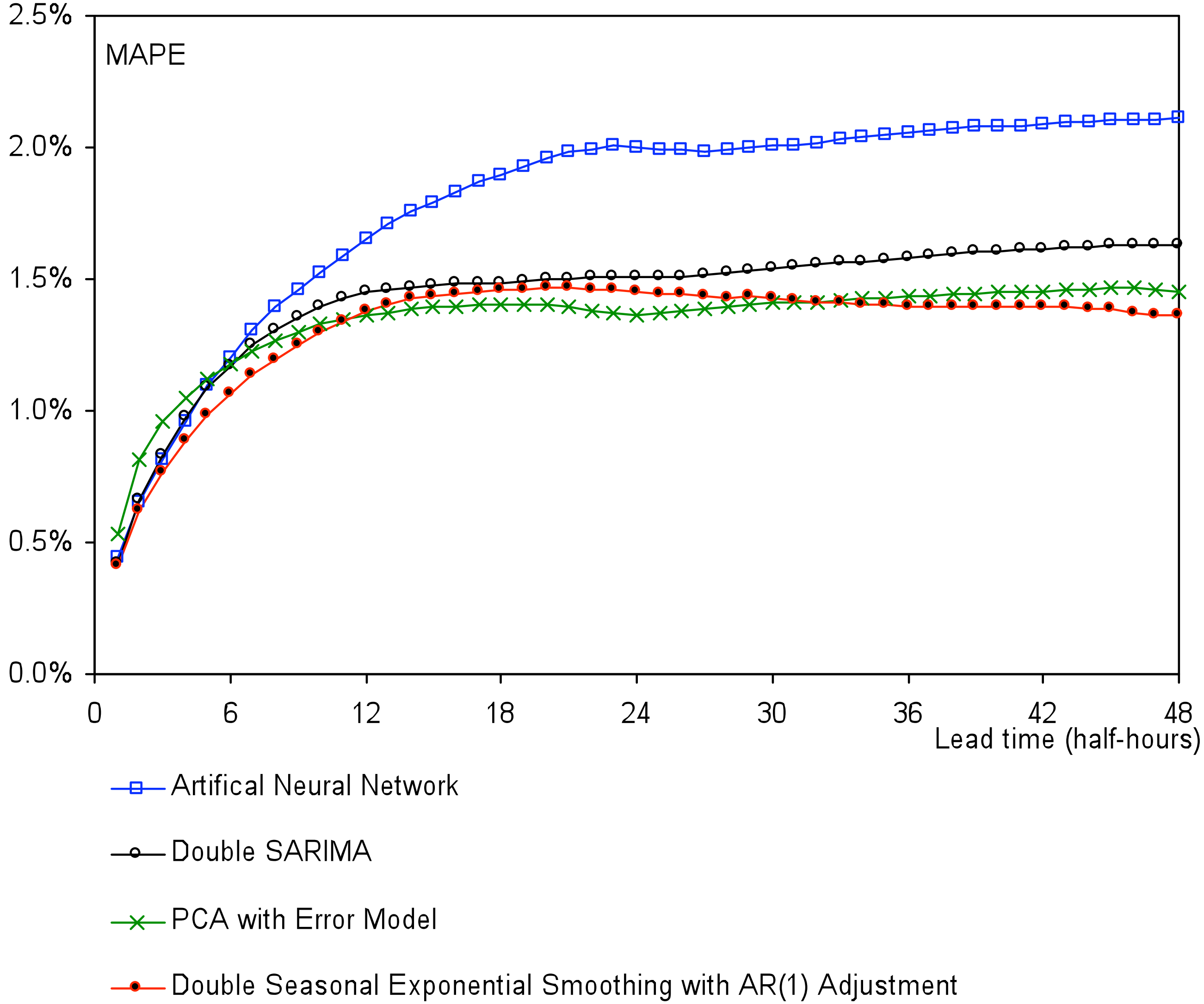
- Accuracy improved by a adjusting 1st-order autocorrelation:
  - AR(1) model, $e_t = \lambda e_{t-1} + \xi_t$, is fitted to the 1-step-ahead in-sample errors (residuals), e_t.
  - k-step-ahead forecasts from forecast origin τ are then modified by adding the term $\lambda^k e_\tau$.

Parameters estimated (20 weeks)

	Level α	Trend γ	Within-day seasonality δ	Within-week seasonality ω	AR λ
Rio	0.01	0.00	0.09	0.15	0.88
England and Wales	0.02	0.04	0.32	0.15	0.98

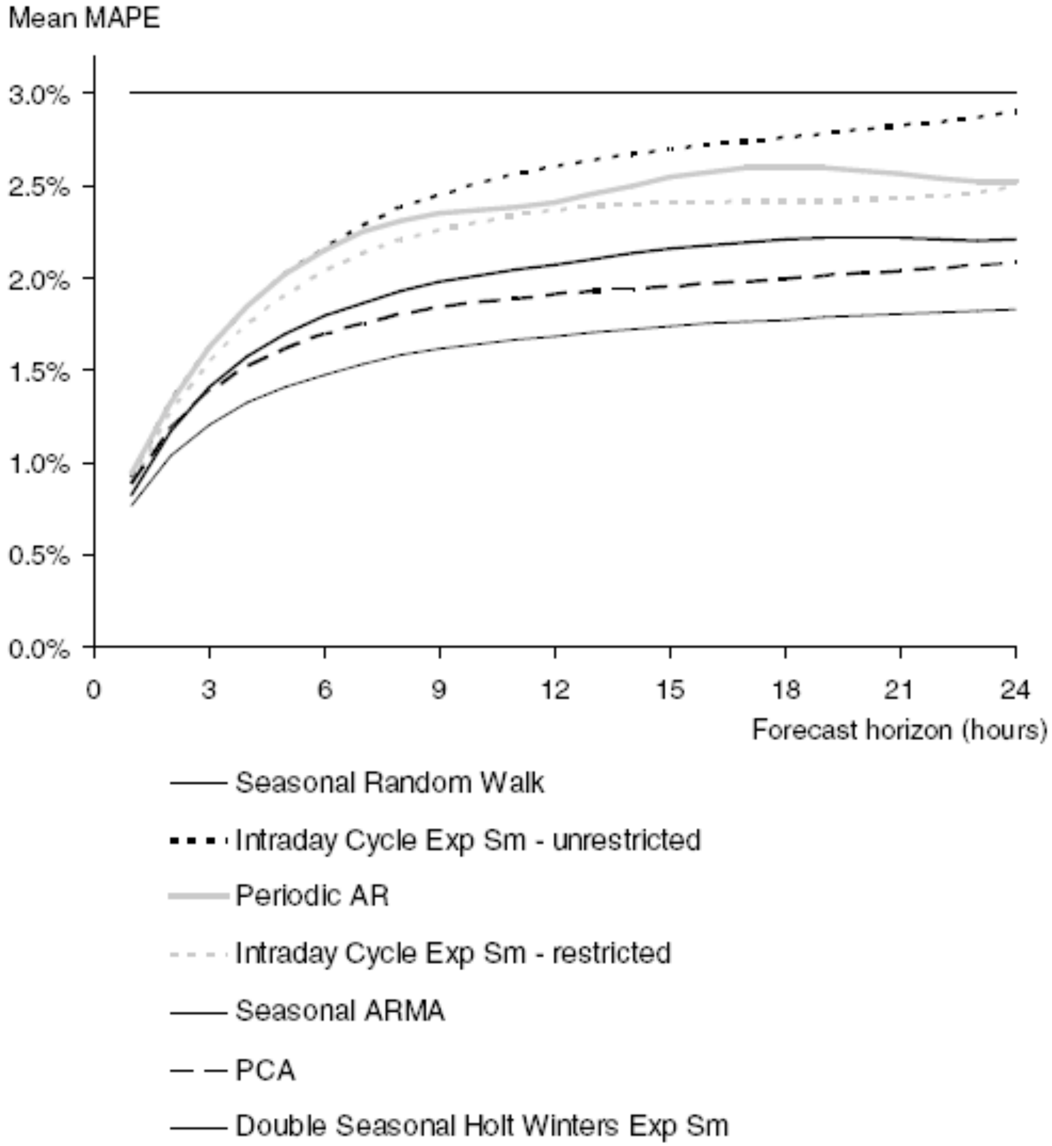
$$S_t = \alpha\,(X_t / (D_{t-S_1} W_{t-S_2})) + (1-\alpha)(S_{t-1} + T_{t-1})$$
$$T_t = \gamma\,(S_t - S_{t-1}) + (1-\gamma)\,T_{t-1}$$
$$D_t = \delta\,(X_t / (S_t W_{t-S_2})) + (1-\delta)\,D_{t-S_1}$$
$$W_t = \omega\,(X_t / (S_t D_{t-S_1})) + (1-\omega)\,W_{t-S_2}$$

Short term energy forecasting



Taylor, McSharry & DeMenezes (2006), International Journal of Forecasting 22: 1-16.

Results for ten EU countries



Taylor & McSharry (2008), IEEE Trans Power Systems 22(4): 2213-2219.

All models are wrong!

- "All models are wrong, but some models are useful" - George P. E. Box
- "Whatever can go wrong, will go wrong" - Murphy's Law (Sod's Law)
- All real-world systems are nonlinear and non-stationary: structural breaks, external shocks
- Any useful model should account for the fact that the unknowable is likely to happen

Netflix competition

- Netflix offered a $1 million prize to anyone who could significantly improve its movie recommendation system Cinematch (with an RMSE of 0.9525) by 10%
- The winning team, "BellKor's Pragmatic Chaos", a group of 7 individuals, achieved 10.06%
- The runners-up, "Ensemble", formed from a collection of 28 teams, achieved 10.06%
- A 50/50 blend of the two would have achieved 10.19%

Ensemble modelling

- Combining knowledge and data-driven approaches
- No single perfect model exists
- Model selection often depends on particular realisation of time series or database available
- Identification of multiple predictive signals
- Ensembles provide a means of pooling predictive information

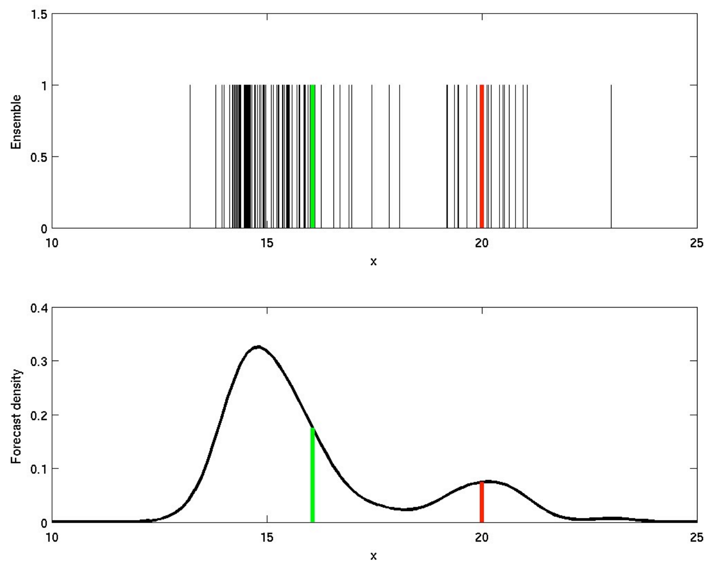
Telemedicine monitoring of Parkinson’s disease

- Intel Case Studentship investigating the novel mapping of voice signals to Unified Parkinson’s Disease Rating Scale (UPDRS)
- Accurate telemonitoring of PD progression (2.5 % error)
- Objective, machine learning implementations
- Remote monitoring, less time-consuming
- Technology enables large scale clinical trials and facilitates design and test of novel drug treatments

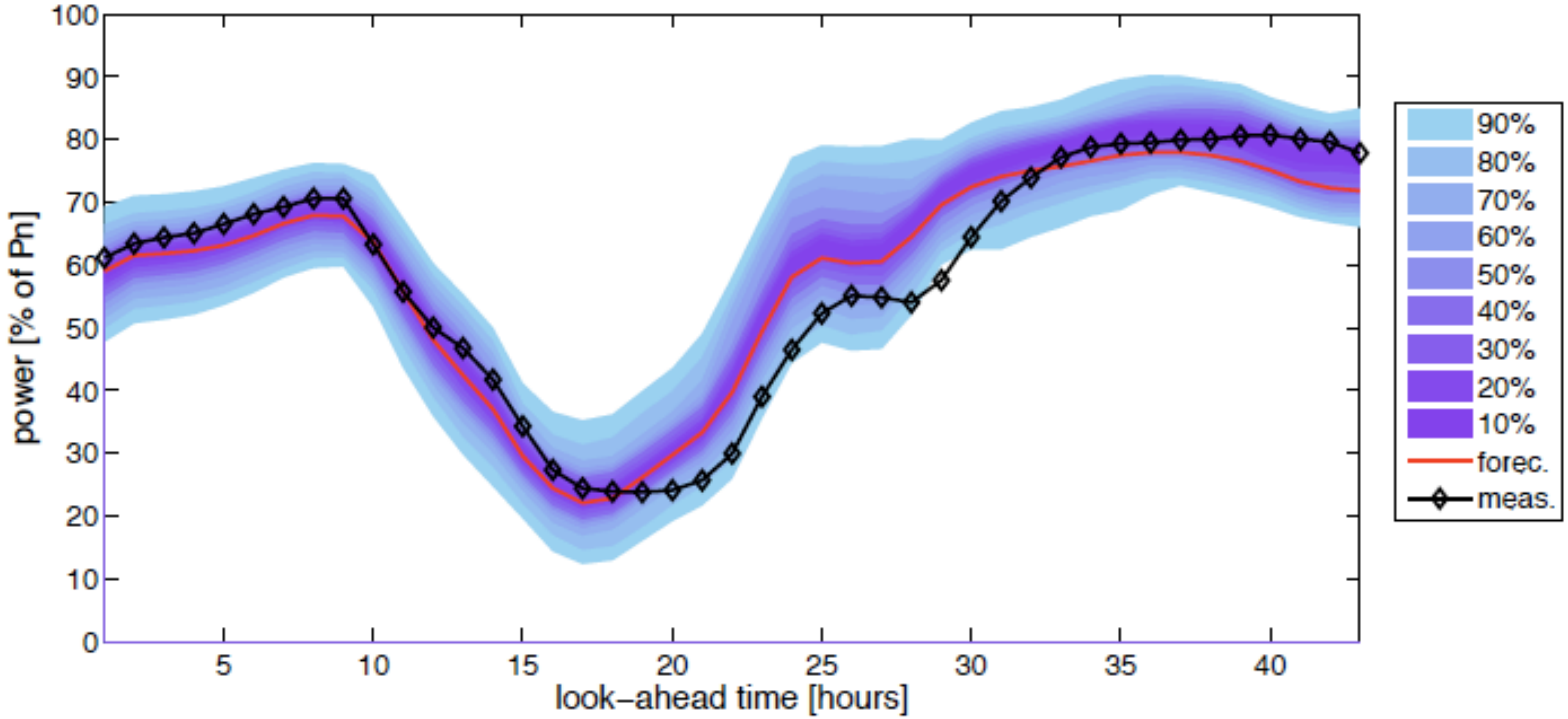
Tsanas, Little, McSharry & Ramig (2009), IEEE Trans Biomed Eng (accepted)

Point and density forecasts

- We need density forecasts in order to account for uncertainty (observational, parametrical, model error)
- Decision-makers like certainty but point forecasts generally convey over-confidence
- Reliable and skillful density forecasts can provide improved point forecasts
- Point forecasts could be supplemented using prediction intervals or risk indices



Probabilistic forecast of wind power in Denmark

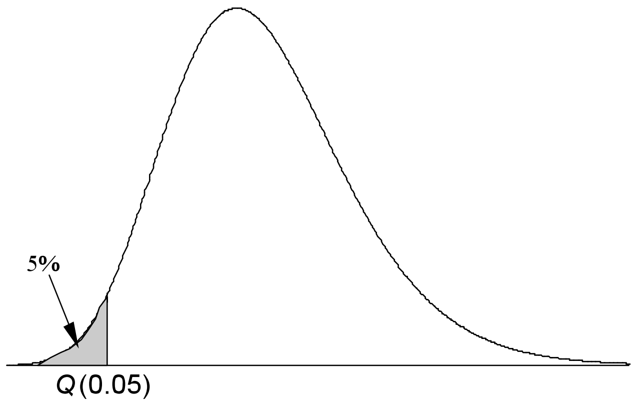


Evaluating probabilistic forecasts

- Value: we can assess a forecast based on its value (economic) for improving decision-making
- Quality (statistical assessment)
  - Reliability: probabilistic correctness of the forecasts
  - Sharpness: ability to concentrate distribution
  - Resolution: accuracy conditional on explanatory variables
  - Skill: overall assessment of quality

Pinson, McSharry & Madsen (2010), Quarterly Journal of the Royal Meteorological Society (online).

Dynamical Risk Estimation



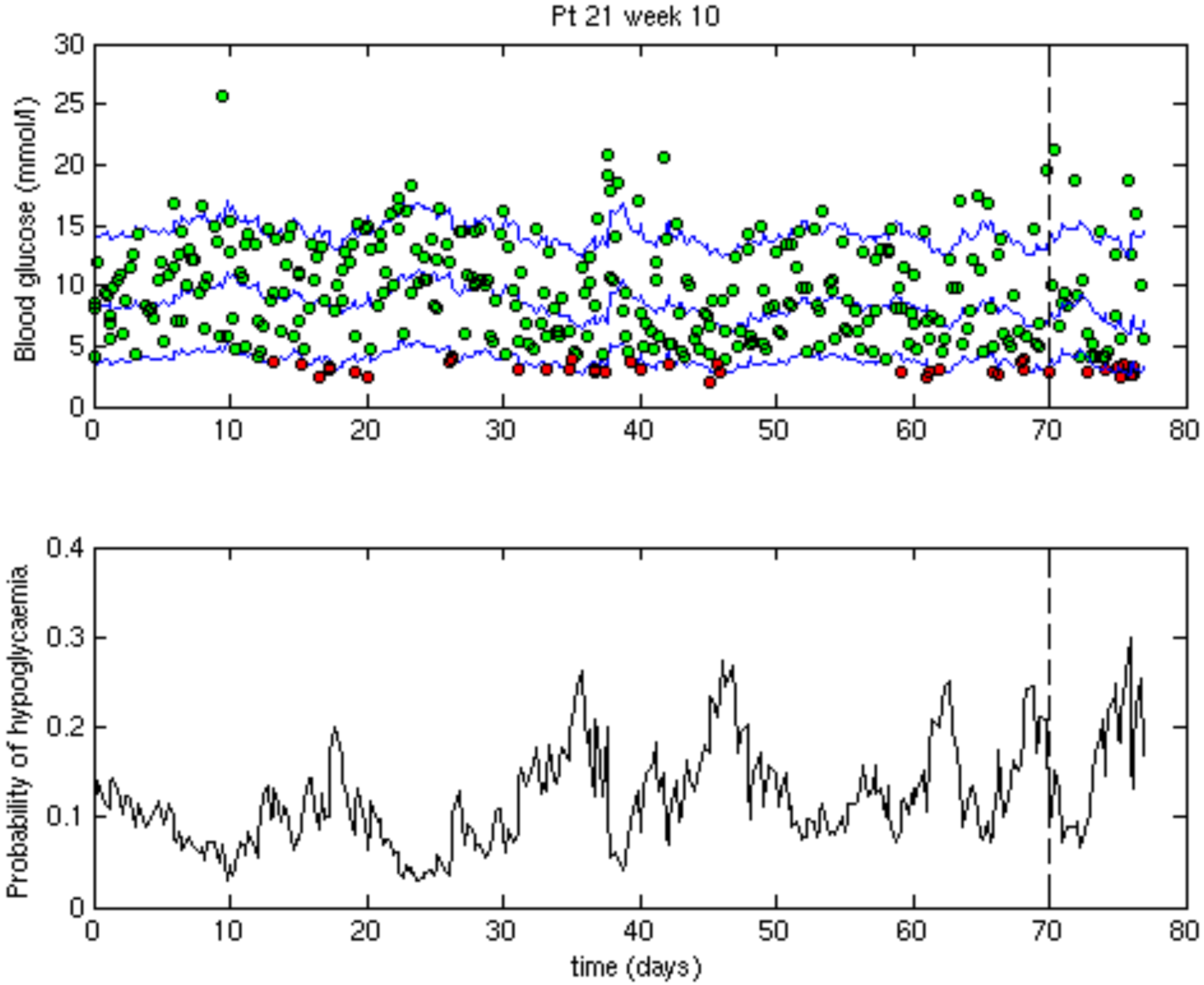
- Conditional Autoregressive Risk Estimation

$$Q_t(\theta) = \omega + \alpha\, Q_{t-1}(\theta) + \beta x_{t-1}$$

- Quantile Regression objective function:

$$\min \left[\sum_{t \mid y_t \geq Q_t} \theta \mid y_t - Q_t \mid + \sum_{t \mid y_t < Q_t} (1-\theta) \mid y_t - Q_t \mid \right]$$

Forecasting risk of hypoglycaemia

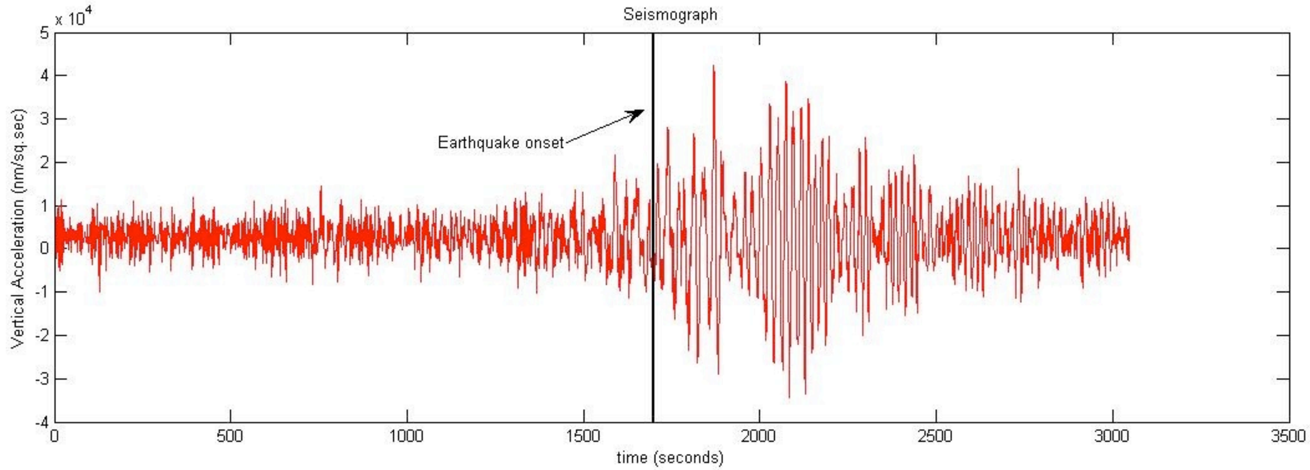


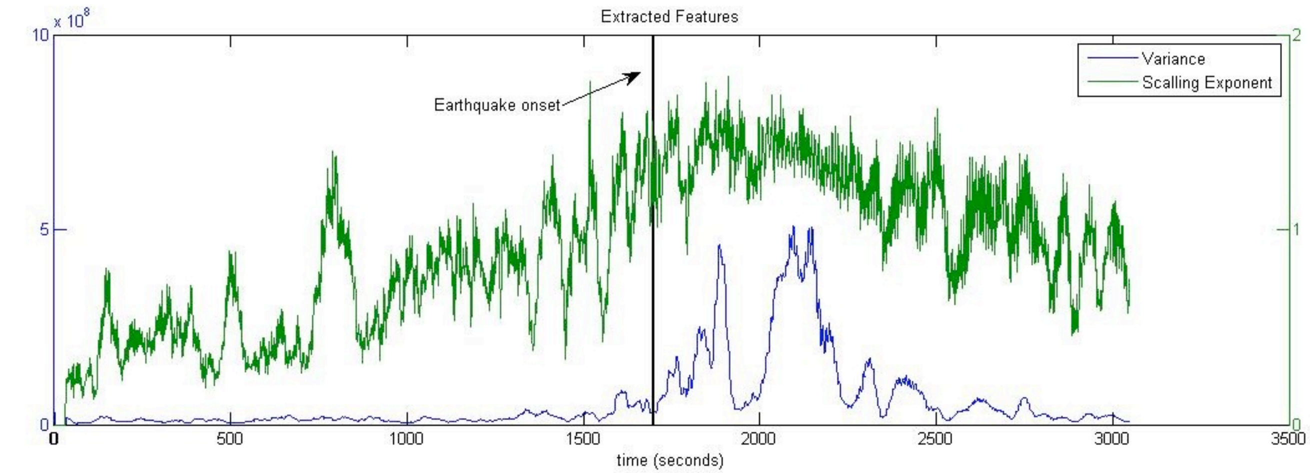
Gibson, McSharry & Tarassenko (2006), Diabetes Care 29.

Multimodel adaptive approach

- All real-world signals are generated by non-stationary processes
- Influences outside of our information set may cause dramatic structural changes
- Construct an ensemble of simple models which describe the data for short time periods
- Combine the models to produce a density forecast

Earthquake prediction (Kobe)





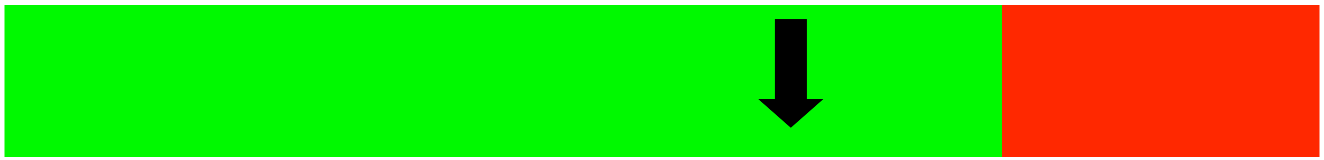
Combining using adaptive weights

- Dynamic weighting scheme, whereby the combination weights for each feature changes with time
- Weights based on each feature's forecasting or classification performance relative to other features
- Measure and track performance using exponentially weighted skill score (EW-SS)

Risk communication

Continuous scale, with no pre-defined threshold:

Quantised scale with pre-defined threshold:



Low risk High risk

Conclusions

- Increasing availability of data and computational power is challenging the decision-making of human experts
- Adoption of quantitative decisions varies greatly with business sector (healthcare/energy/insurance/finance)
- Modelling simplifications should be transparent
- Uncertainty at different stages during the modelling process will feed through to the final output
- Ensemble modelling provides an approach for addressing uncertainty for decision-making
- Model combination through adaptive weights is appropriate for data streaming applications