17PESGM2648

Big Data for Integrated Energy Systems

17PESGM2648

Data driven load models

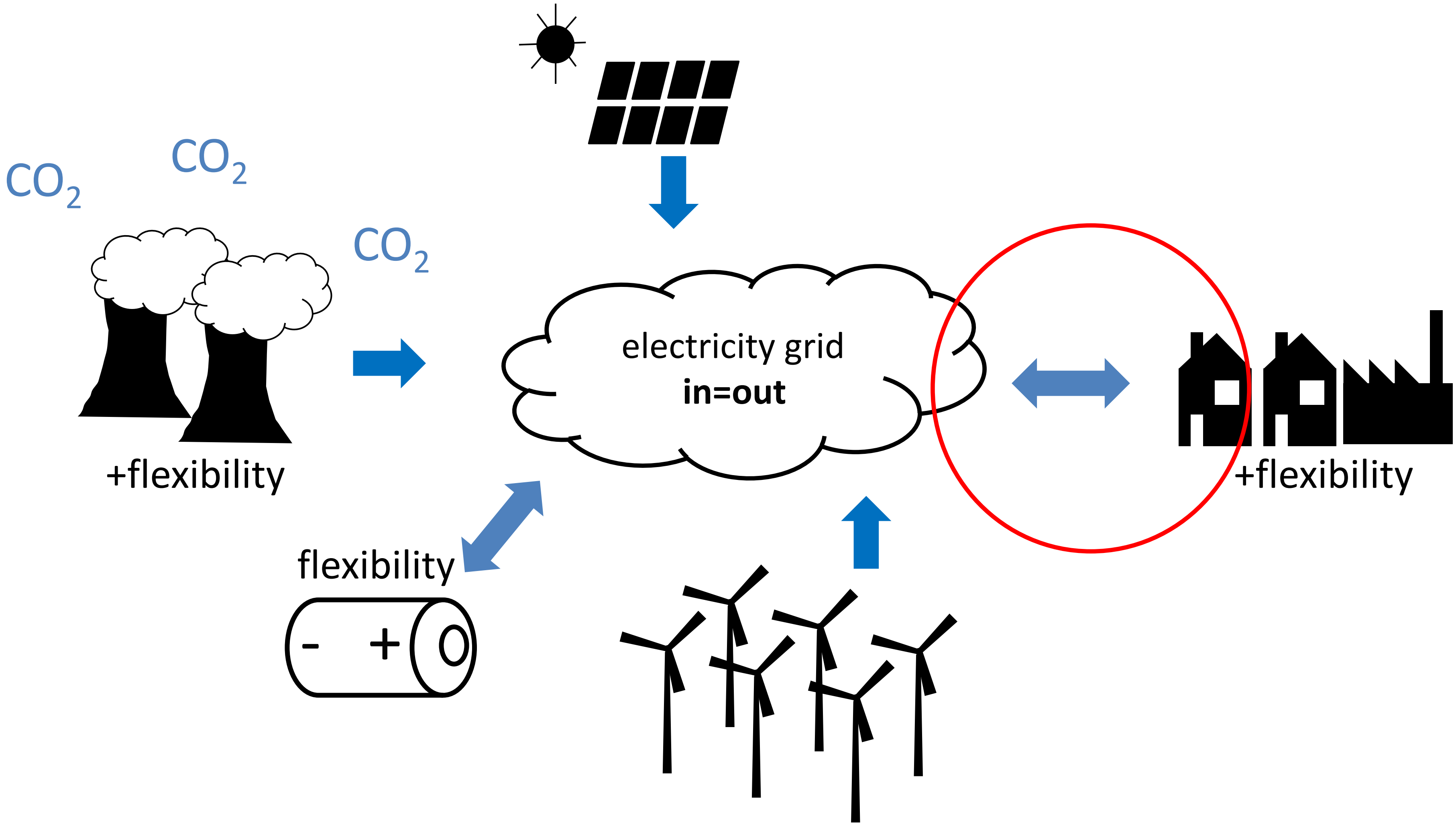
scaling up dimensionality and understanding using growing data and computational tools

Simon Tindemans, **James Schofield**, Miao Wei, **Mingyang Sun, Ioannis Konstantelos**, Goran Strbac, Samir Issad (RTE), Patrick Panciatici (RTE)

Imperial College London

IEEE PES Power & Energy Society

IEEE



IEEE PES Power & Energy Society
IEEE

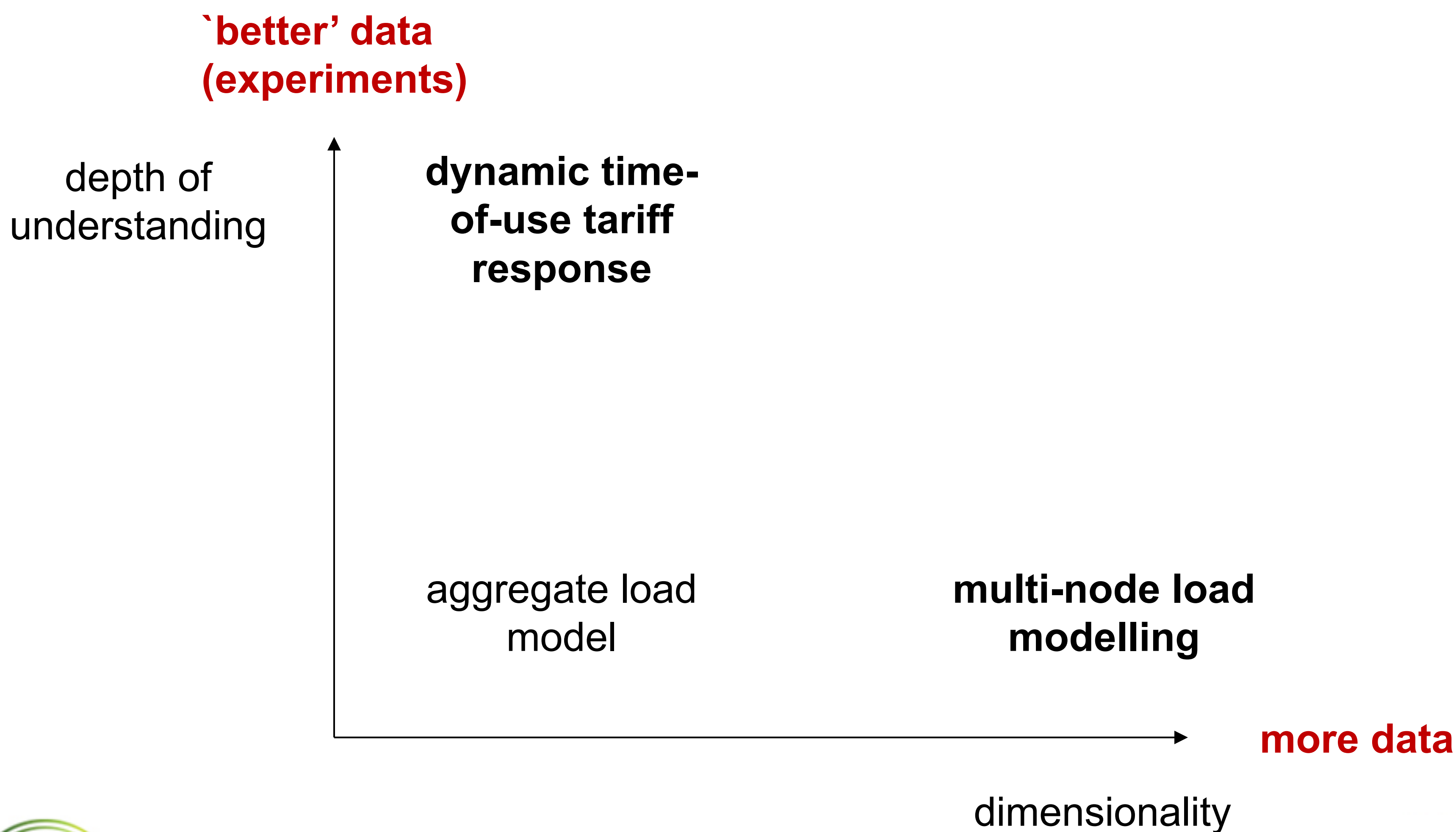
Tighter integration of demand

- Tighter margins at all system levels

- Scarcity of *bottom up* load models
  - Nonexistent (e.g. humans)
  - Not shared (e.g. commercially sensitive)

- Data-driven load models
  - Descriptive models: anticipation
  - Predictive models: control

Load model space



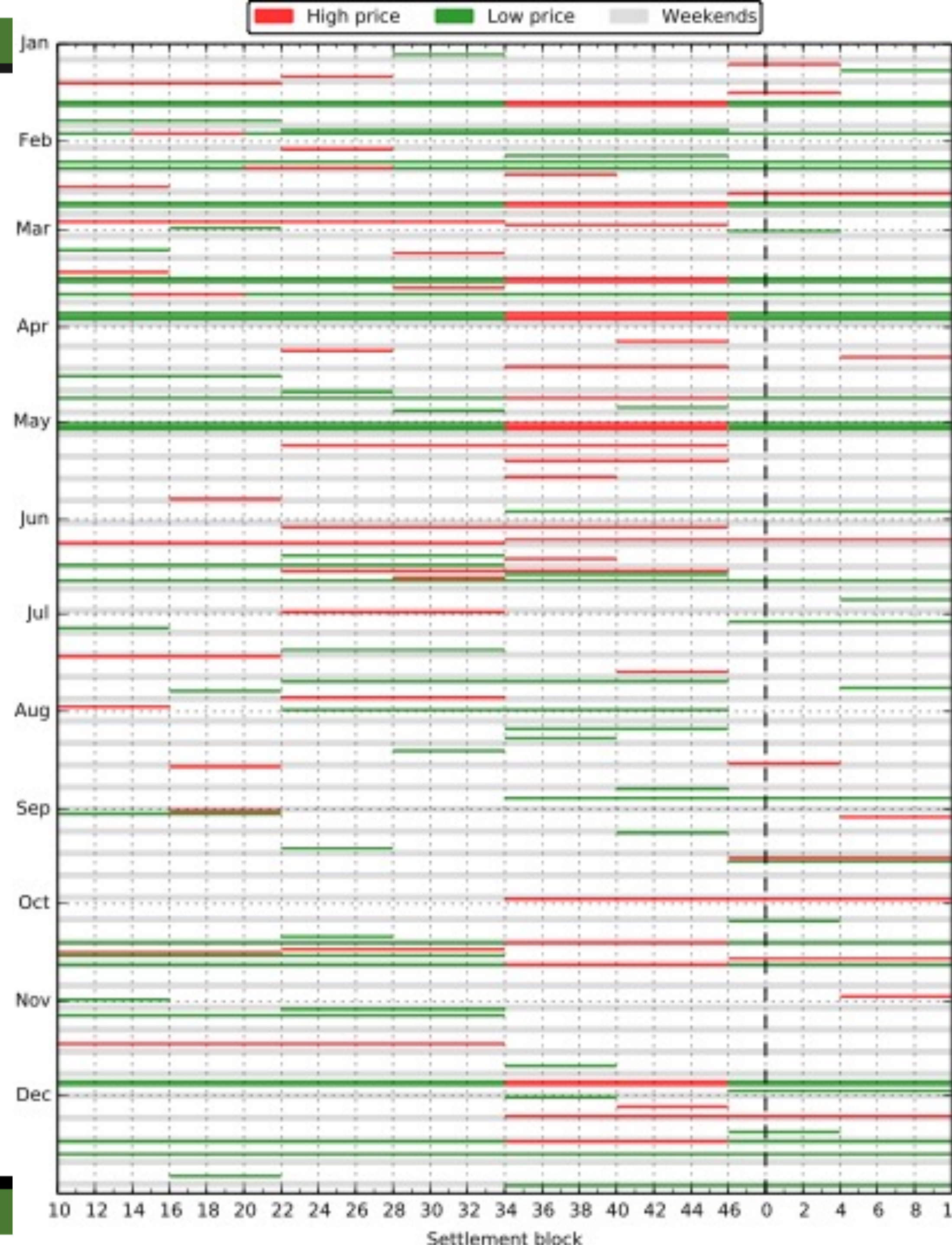
EXTRACTING KNOWLEDGE FROM SMART METER DATA

Low Carbon London

- UK's first **dynamic time-of-use tariff** demand response trial (2013)
- 5536 households with smart meters
- 1119 households took part in dynamic time of use trial
  - Day ahead notification of prices via SMS and in-home displays
  - Three price levels
    - Default: £0.1176/kWh
    - Low: £0.0399/kWh
    - High: £0.672/kWh

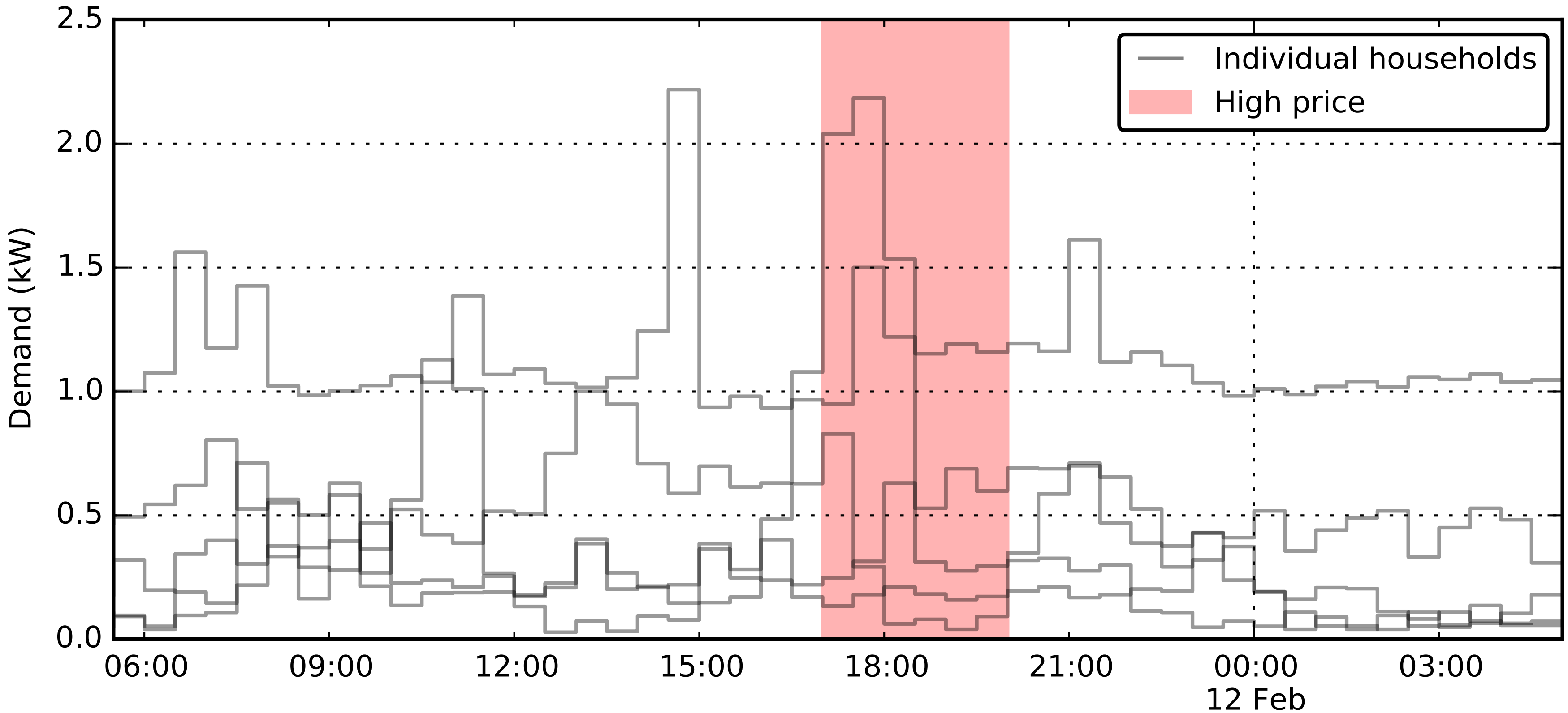
- **93 supply following events**
- 45 high price events (3-12 hours)
- 48 low price events (3-24 hours)

- **13 constraint management events**
- high price, flanked by low prices
- primarily targeted at evening peaks
- 1-3 consecutive days (21 days in total)



IEEE PES Power & Energy Society®

Measured response to events



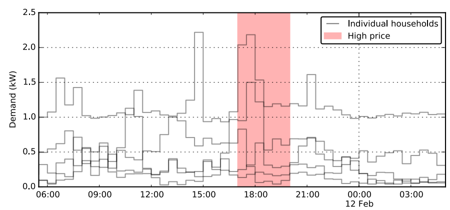
Dataset can be downloaded from UK Data Service
www.ukdataservice.ac.uk ; search for "Low Carbon London"

Extracting knowledge through aggregation

single event

many events

single household



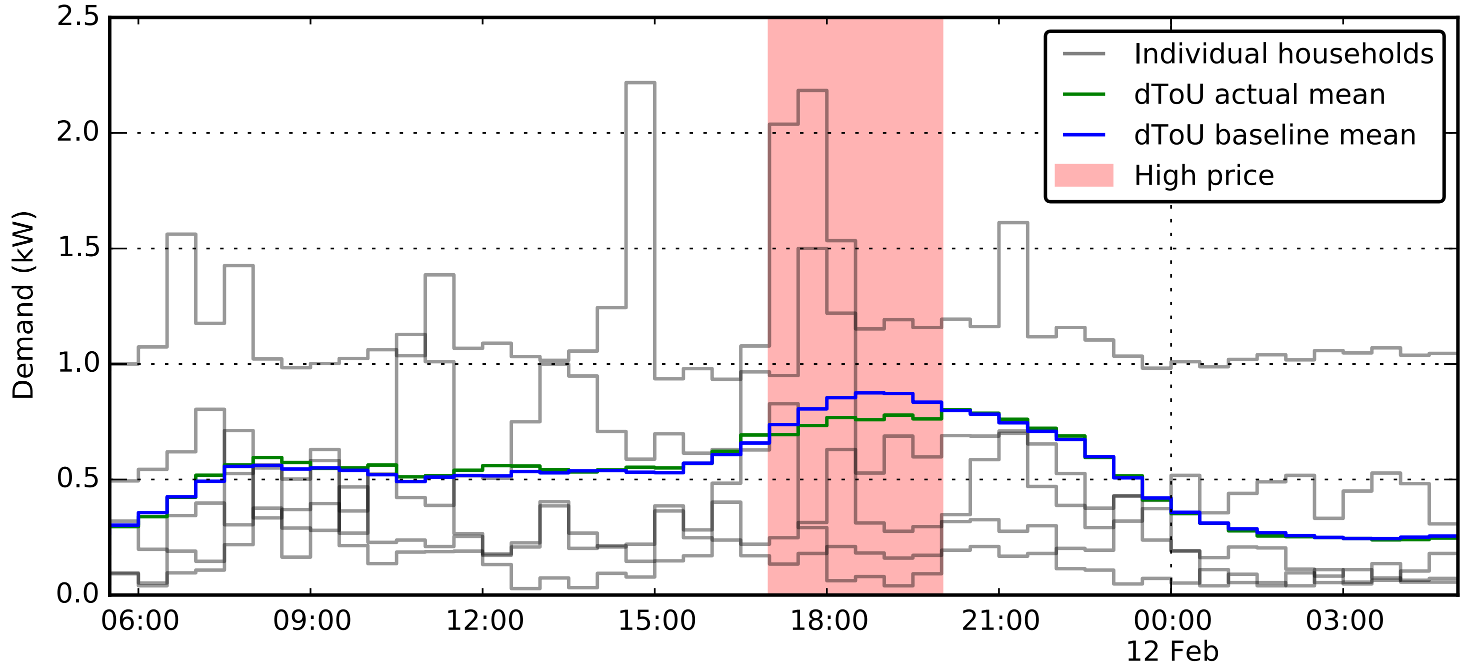
Topic 2:
Analysis of household responsiveness

many households

Topic 1:
Analysis of aggregate response

***Useful* knowledge includes estimate of uncertainty!**

Baselines to measure responsiveness



Construct a linear regression model for the baseline, trained on non-event days.

$$B_t = \sum_{w=1}^{W} \left(\alpha_w d_{w,t} + \beta_w A^{\text{non-tou}}{}_t d_{w,t}\right) + \gamma\, t + \delta T_t$$

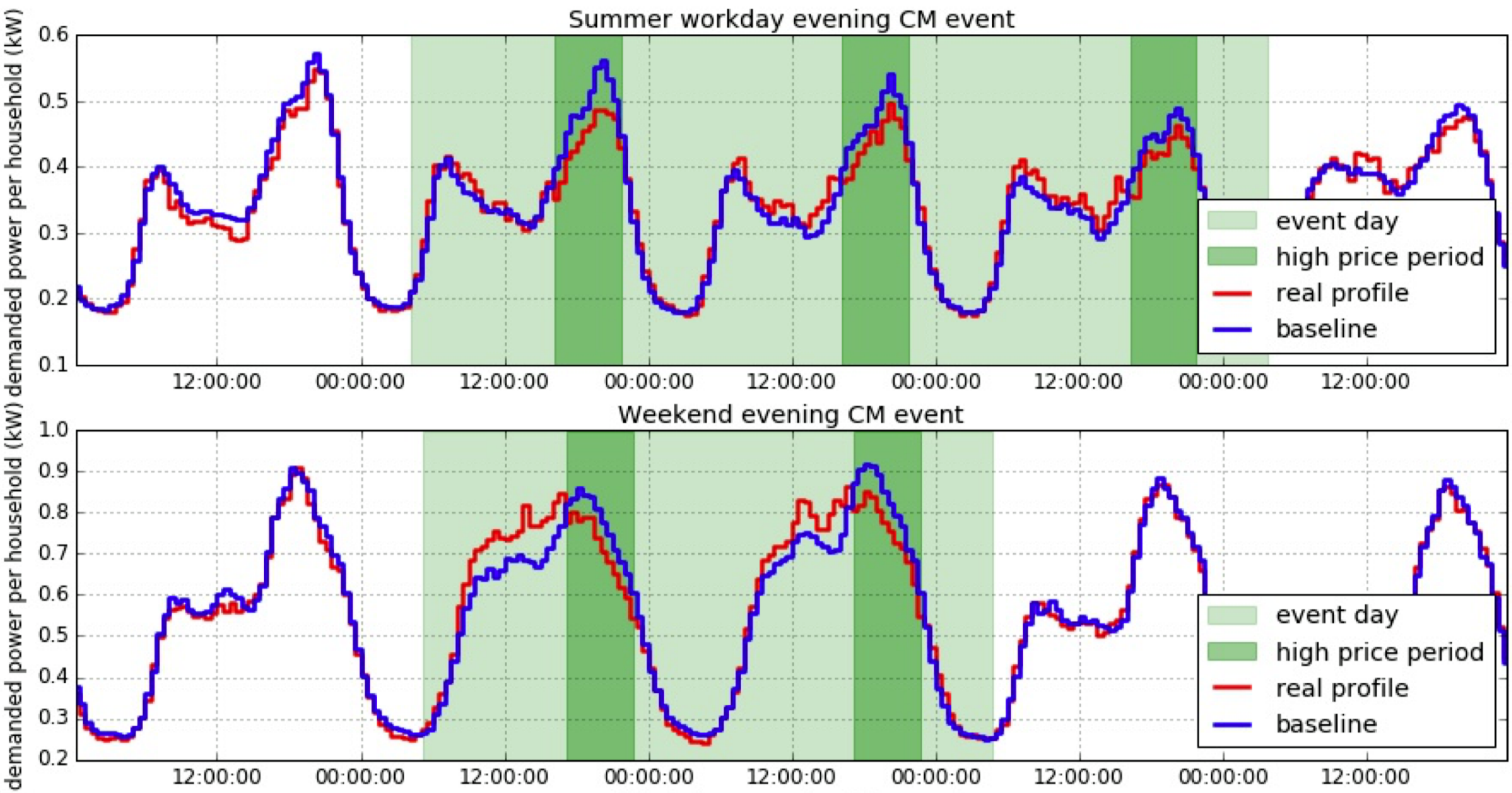
weekly profile

weekly coupling to non-ToU group consumption

trend line

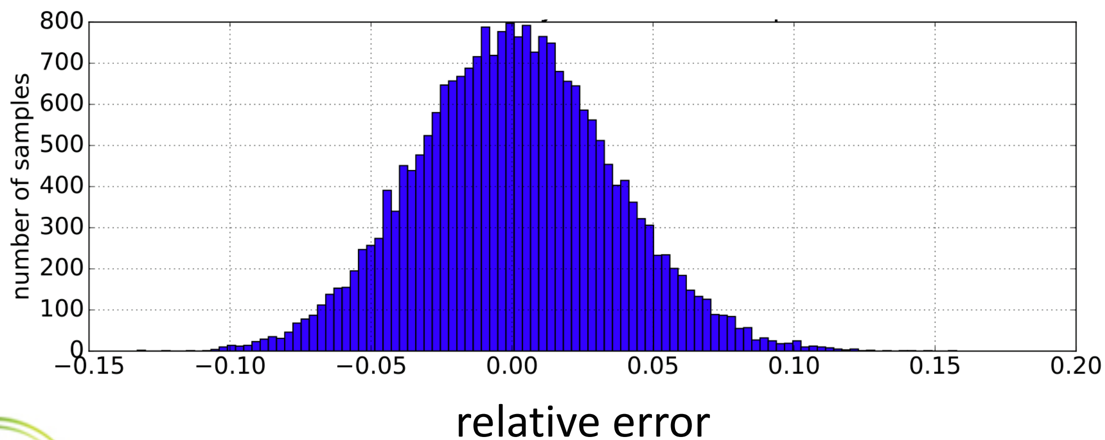
temperature factor

Example of measured response



How good is the baseline model?

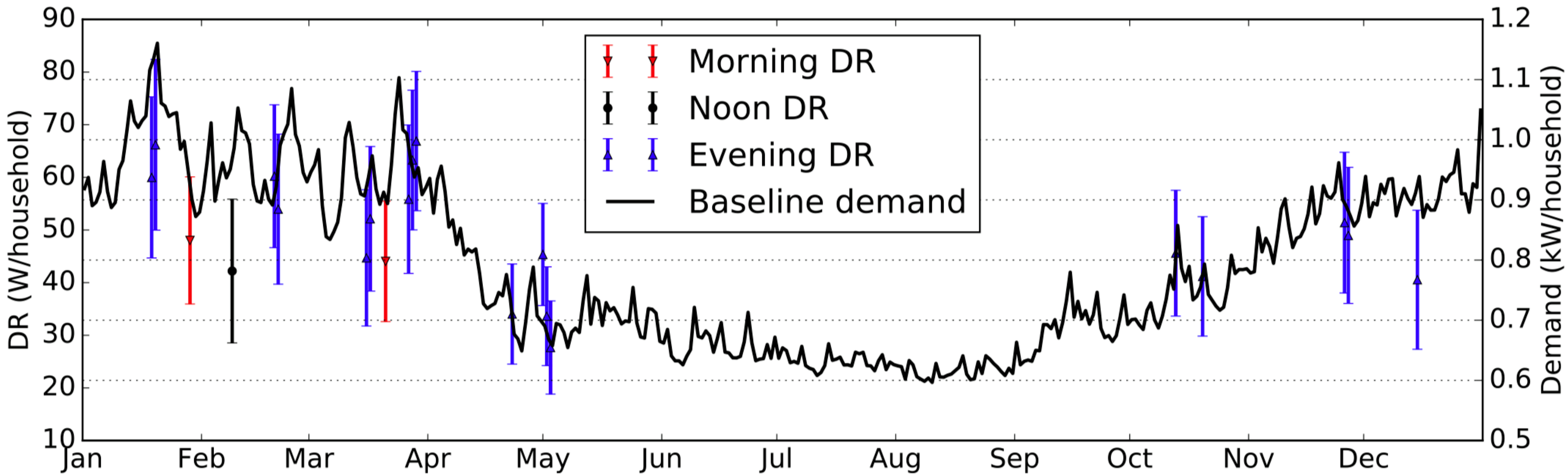
- Bootstrap procedure to select training days
- Train a baseline model for each resampled data set.
- Compute the average out-of-bag error for each 30min settlement block.



DR block	St Dev
30 mins	3.5%
3 hours	2.5%
6 hours	2.0%

baseline error model!

Overview of peak shaving events



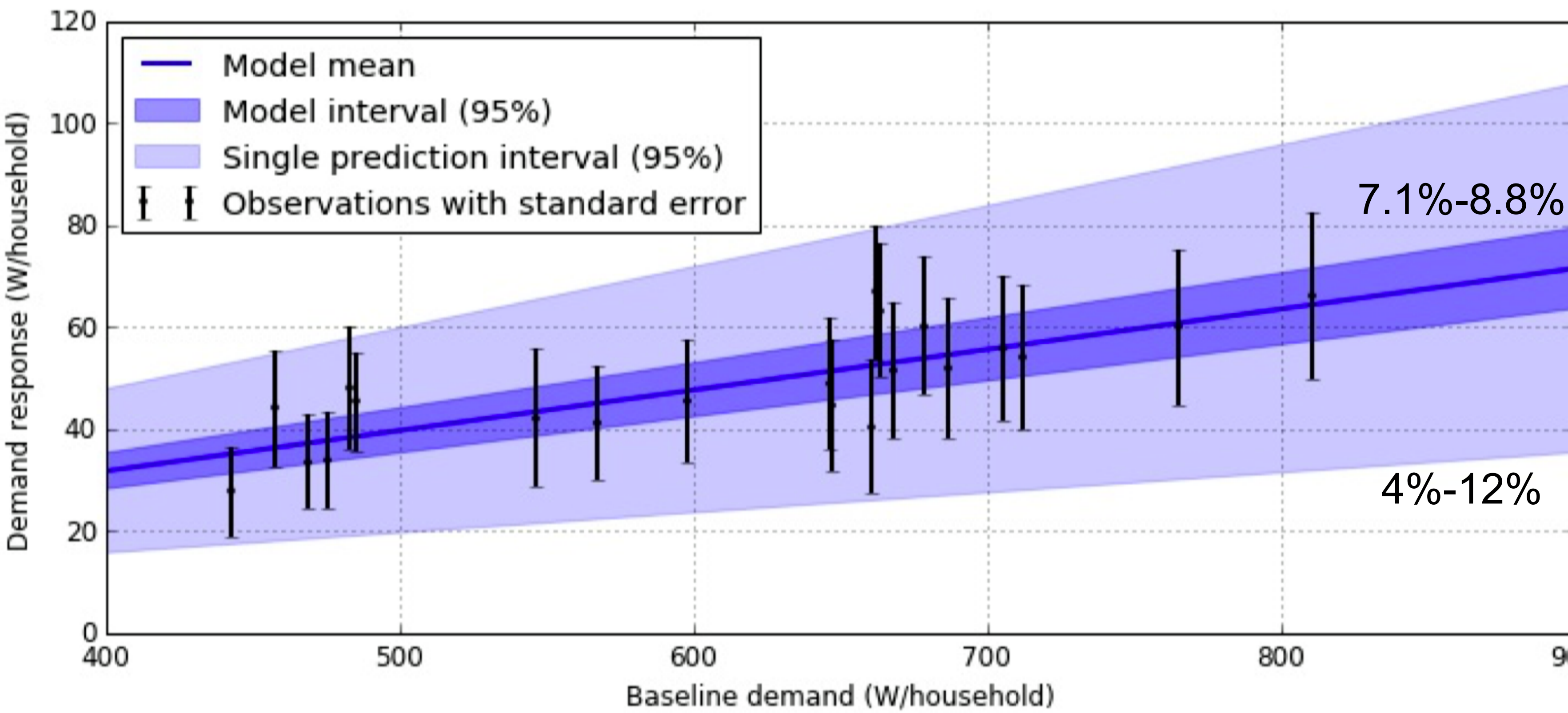
21 events

Toward predictive use of models

Simplest consistent model:

$$R_{CM}^{demand} = -0.079 \times [\text{baseline demand}] + (\text{random variation})$$

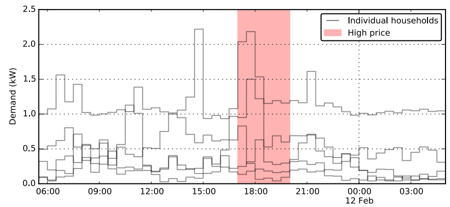
Uncertainty in parameters + baseline variability



single event

many events

single household



Topic 2:
Analysis of household responsiveness

many households

Topic 1:
Analysis of aggregate response

Identifying ‘responsive’ households

Naive approach: Change in bills
Compare actual bill with hypothetical bill on a flat tariff

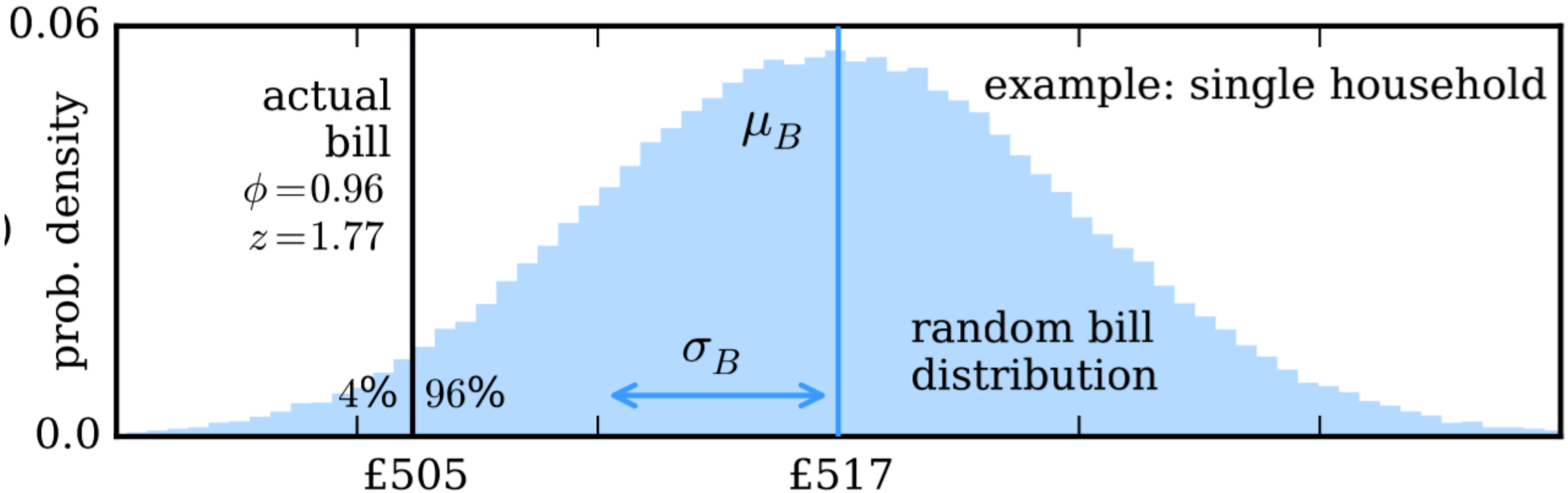
Proposed approach: resampling

1. Compute the actual bill b^* using the actual price signal p_t and consumption c_t:
$$b^* = \sum_{t=1}^{T} p_t c_t$$
2. Generate randomised bills tariffs by permuting daily price signals
$$B = \sum_{t=1}^{T} p_{\Pi(t)} c_t$$
3. Compare the true and hypothetical bills

James Schofield, Simon Tindemans, Goran Strbac, arXiv:1605.08078

IEEE PES Power & Energy Society® IEEE

Nonparametric responsiveness measure



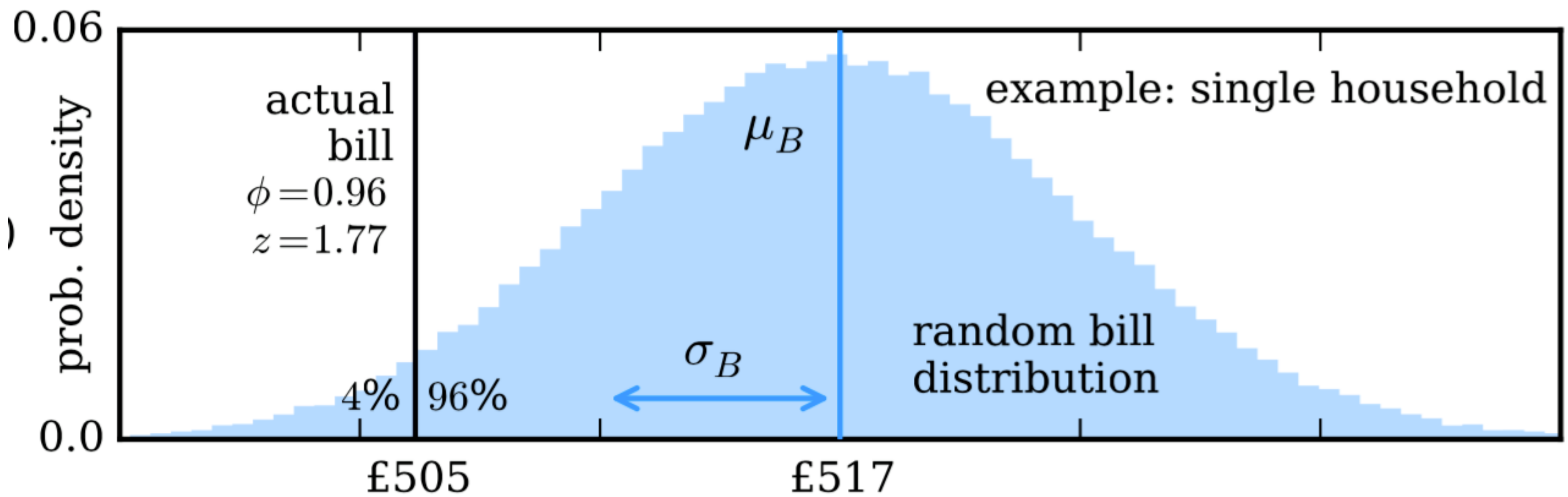
Define a measure of responsiveness:

$$\varphi = \Pr(\mathrm{B} > b^*)$$

B = random variable
b^* = actual bill

B is approximately normal, so φ has an intuitive interpretation as a signal-to-noise measure.

Interpreting per-household responsiveness



What makes a household 'responsive'?

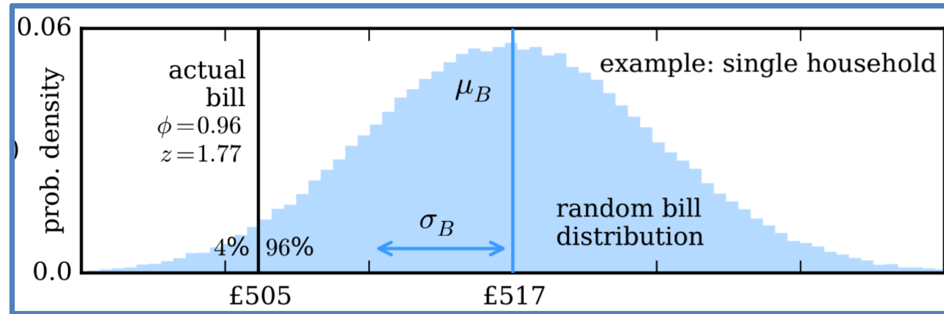
1. Deliberate demand response
2. 'Accidental' demand response ← quantify
3. Price signal bias, relative to the population's consumption pattern ← eliminate

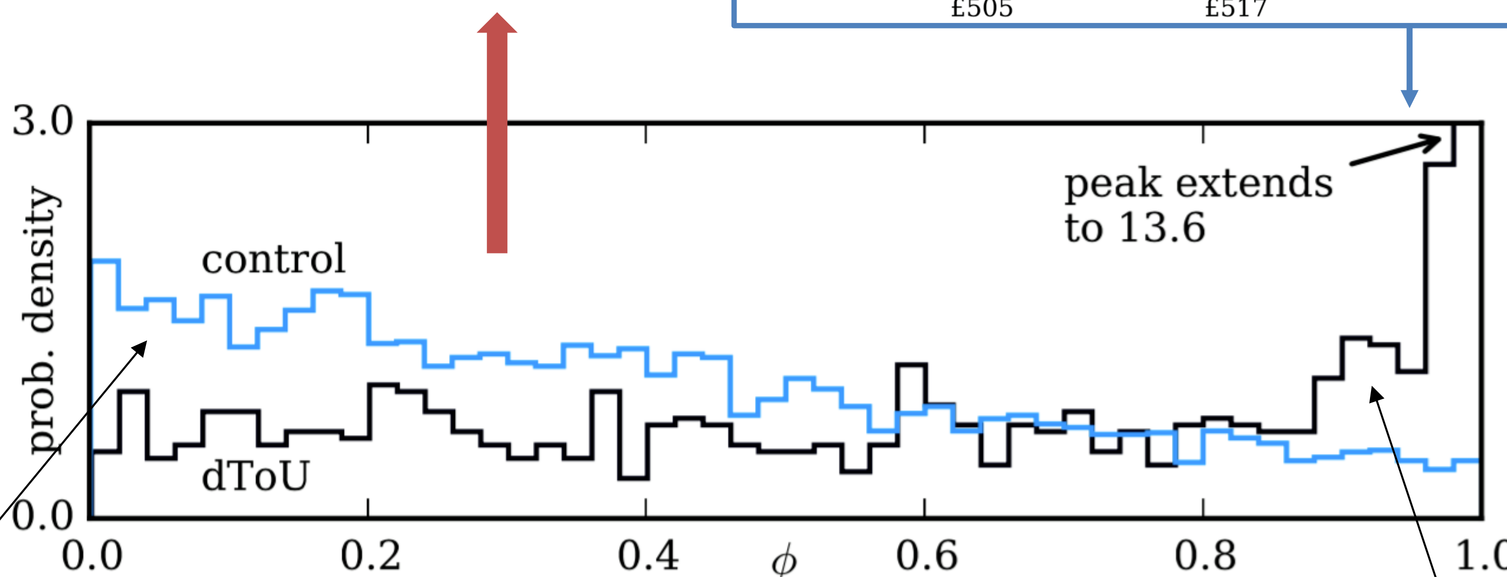
We can dig deeper using data from a control group

Correcting for price signal bias

Reparametrize distribution to correct for price bias

$$\psi = F_{control}(\varphi)$$

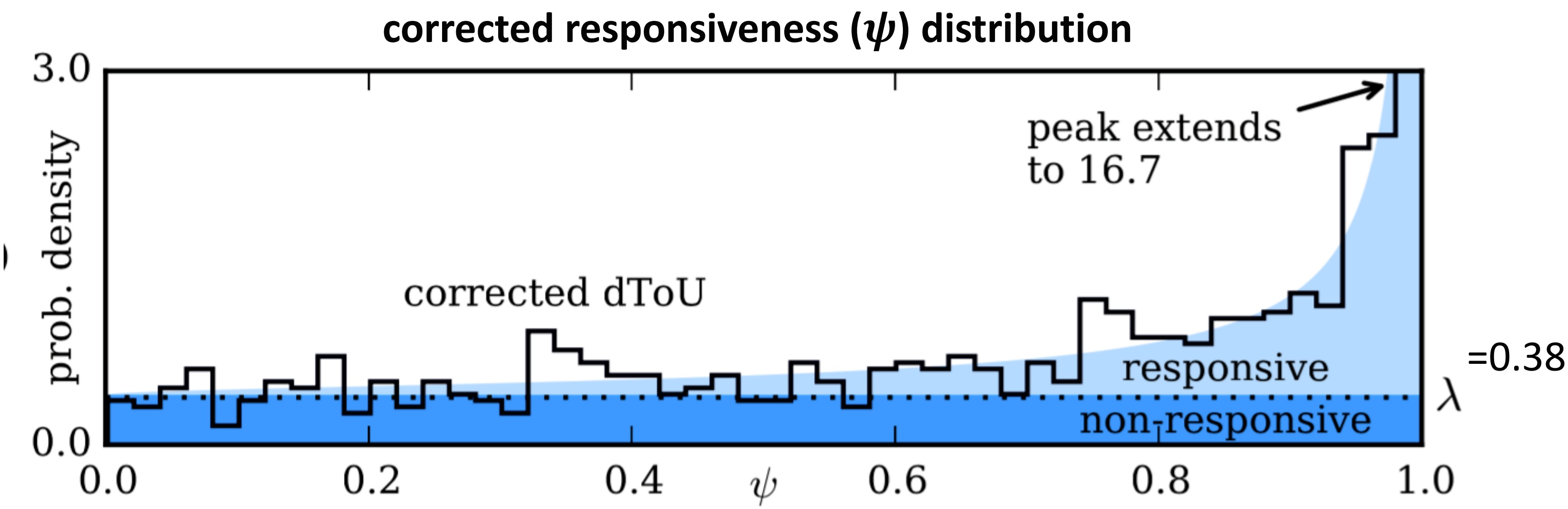




Evidence of price signal bias

Evidence of significant demand response

Quantifying household responsiveness



62% of households are part of a responsive subpopulation

Each household has **probabilistic** measure of responsiveness:

$$\Pr(responsive|\psi) = \frac{f(\psi;\lambda) - \lambda}{f(\psi;\lambda)}$$

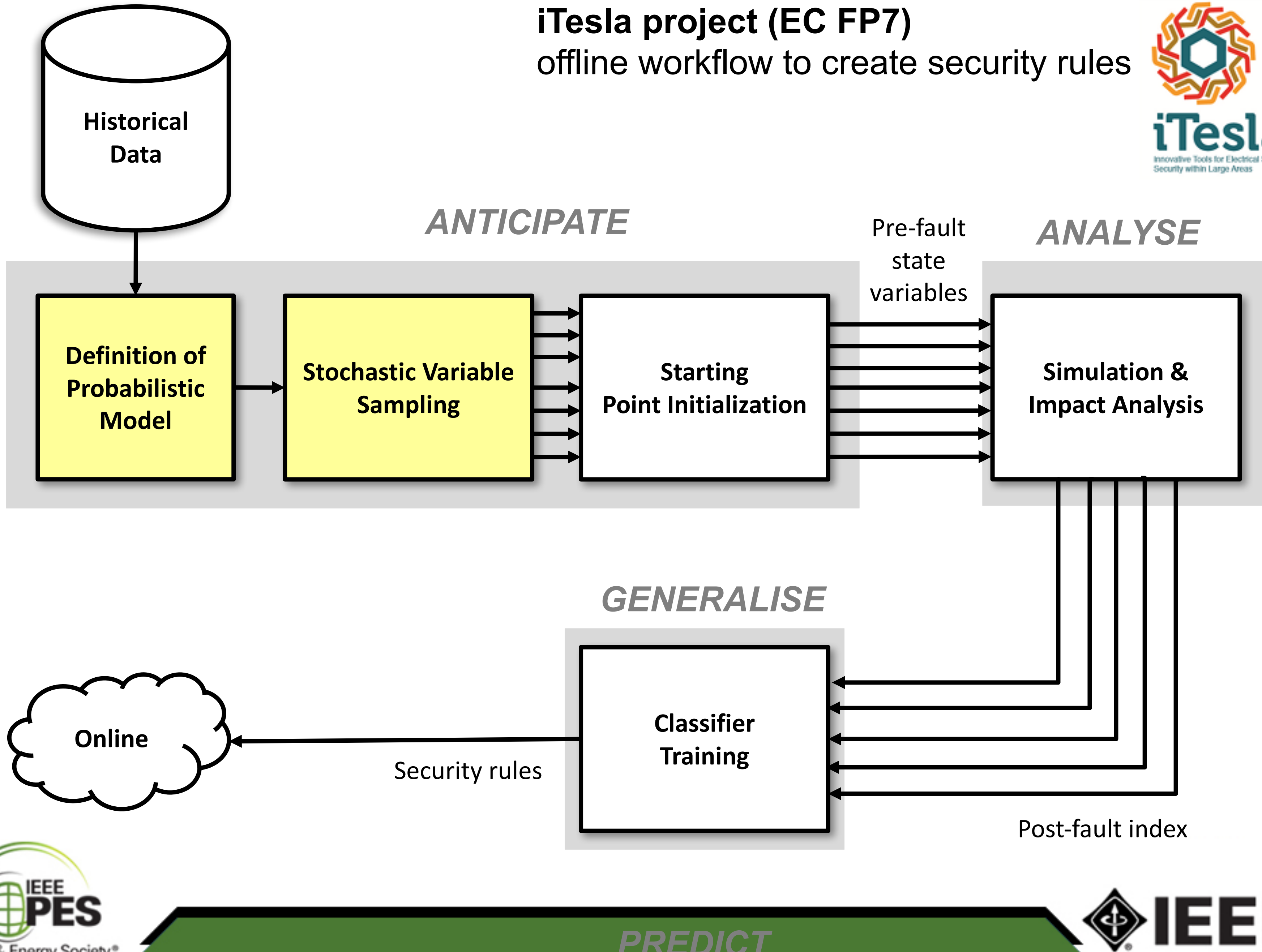
HIGH-DIMENSIONAL LOAD MODELLING FOR MACHINE LEARNING

Use case

- Comprehensive Dynamic Security Assessment requires
  - Time-domain simulations
  - For all credible contingencies
  - For a range of scenarios
  - On a number of timescales

- Use machine learning to construct `proxies' (aka emulators) for simulation outcomes. See e.g.
  - Panciatici, P., Bareux, G. & Wehenkel, L., 2012. Operating in the Fog: Security Management Under Uncertainty. *IEEE Power and Energy Magazine*, 10(5)

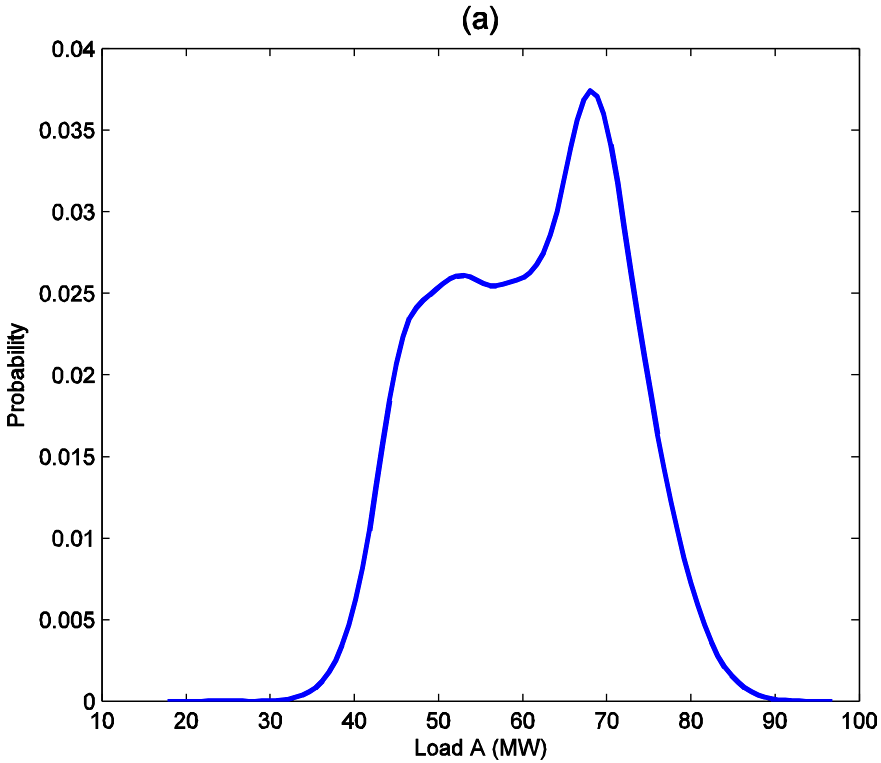
iTesla project (EC FP7)
offline workflow to create security rules



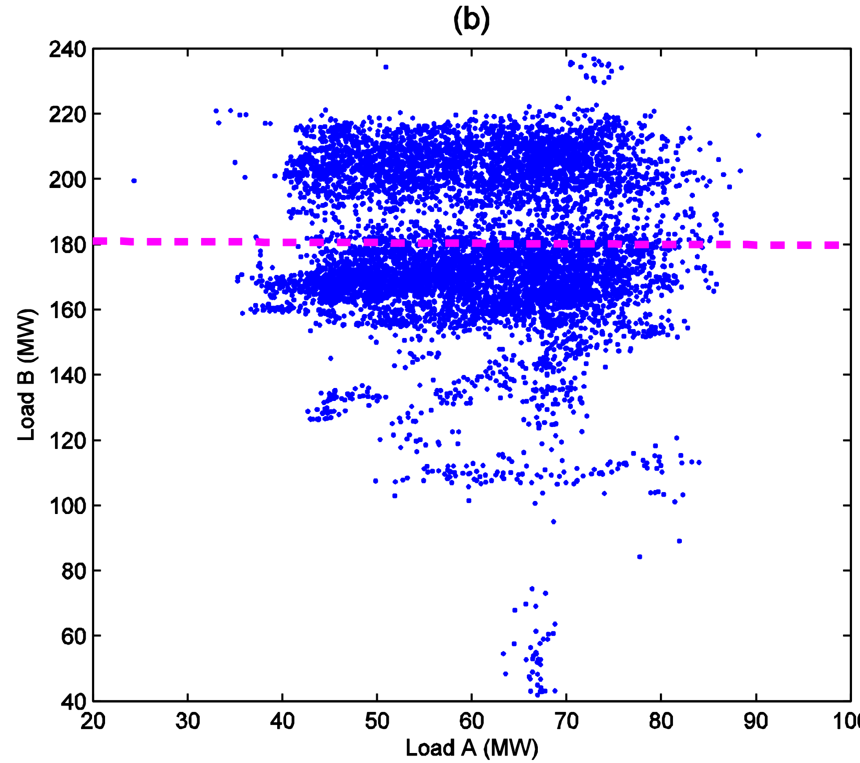
Requirements for probabilistic model

- Model multi-variate stochastic injections/loads (100s)
- Correct sampling of marginal distributions
- Accurately represent dependencies
- Can be used to sample *many* representative points
  - Generate larger sample pool than historical data alone!

Dependency patterns



Marginal probability distribution
15-minute load measurements, 3 months
bus in the region of Nancy, France (2012)



Bivariate empirical distribution
between two load points in the same
region

Copulas: Sklar's theorem

Consider n random variables $\boldsymbol{X} = (X_1, \dots, X_n)$.

Independent variables have a joint probability density function

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$$

Dependent variables have a joint PDF that can be written as:

$$\boldsymbol{f(x_1, \dots, x_n) = c_{1\dots n}(F_1(x_1), \dots, F_n(x_n)) \cdot f_1(x_1) \dots f_n(x_n)}$$

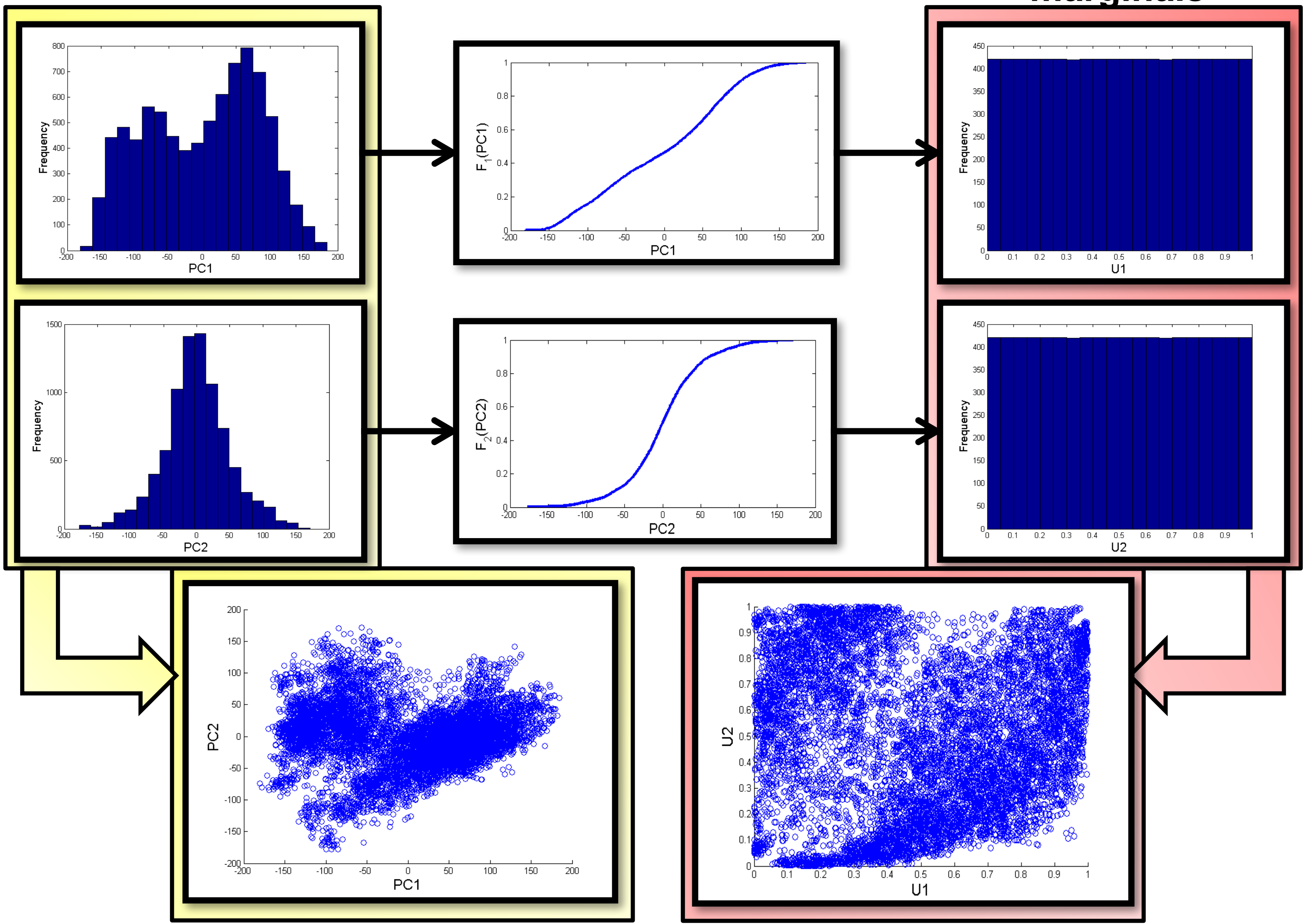
where c is the multivariate copula density.

Copulas allow decoupling of dependency structure and marginals of a multivariate probability function

Decoupling marginals from dependence

marginals

marginals

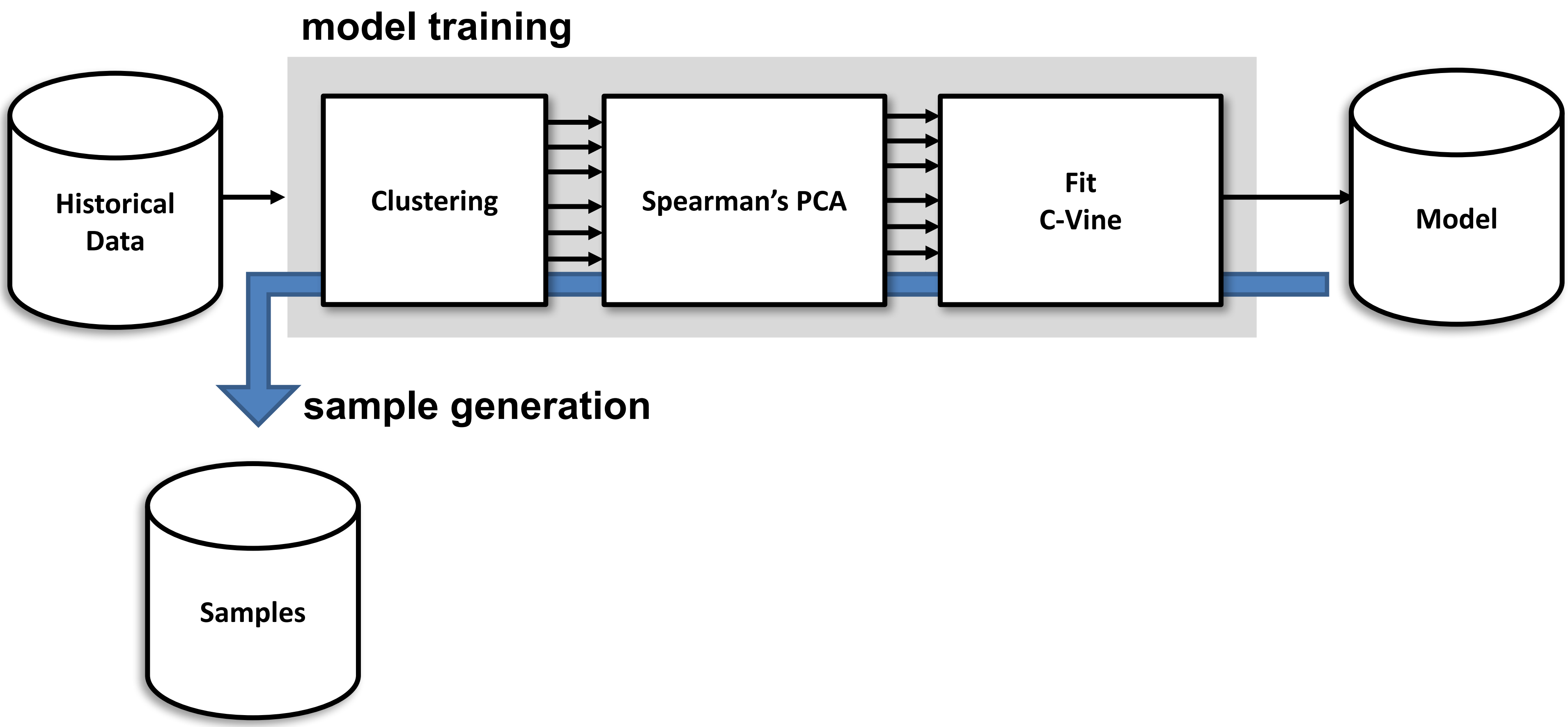


Multivariate copulas

- Wide variety of parametric copulas for bivariate case, but limited options for multivariate distributions
- C-Vine copula (Bedford and Cooke, 2001) uses Pair copula construction (Joe, 1996) to construct multivariate from bivariate copulas
- We truncate the C-Vine to limit impact of dimension

$$f(x_{1:n}) = \prod_{j=1}^{m-1} \prod_{i=1}^{m-j} c_{j+i,j|j-1:1}\left(F_{j+i|j-1:1}, F_{j|j-1:1}\right) \cdot \left(\prod_{k=1}^{m} f_k(x_k)\right) \cdot f_{n:m+1|m:1}(x_{n:m+1}|x_{m:1})$$

The modelling pipeline

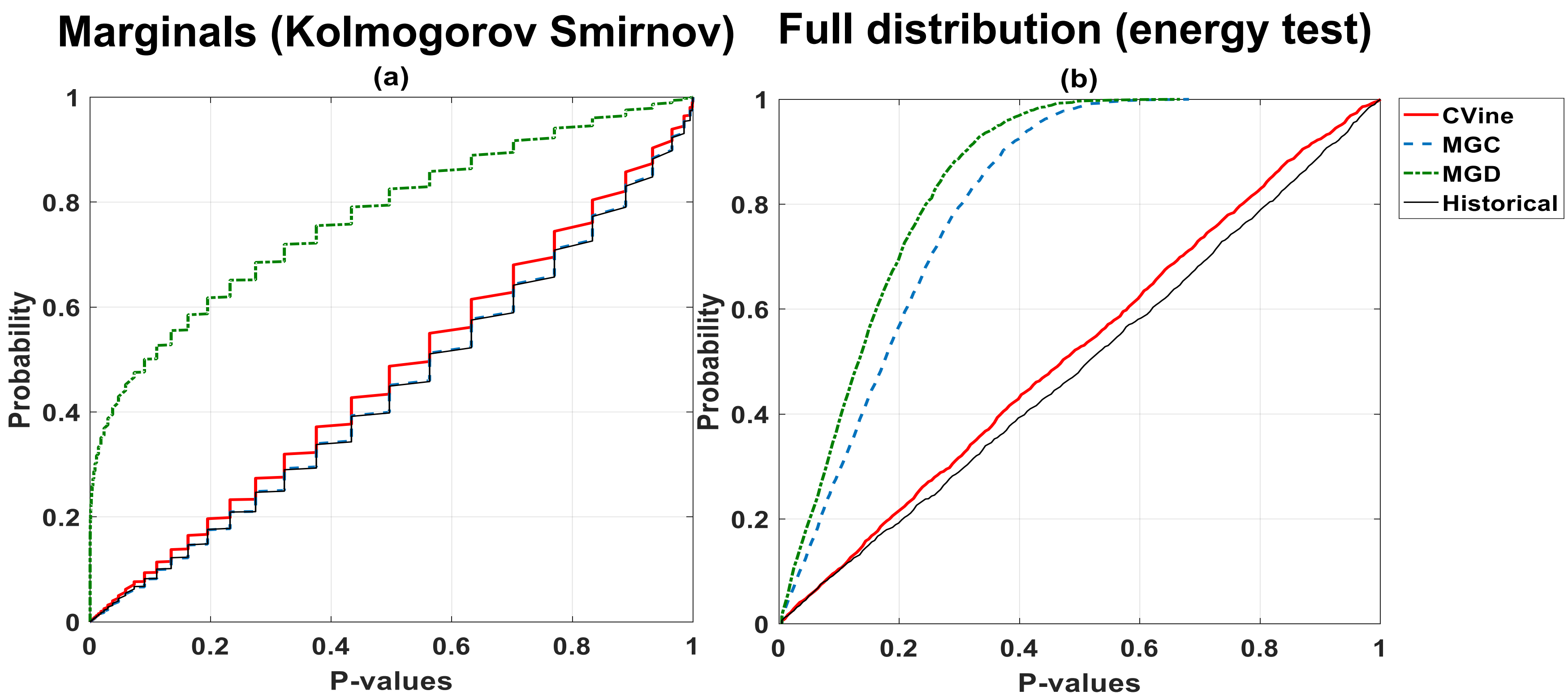


Does it work?

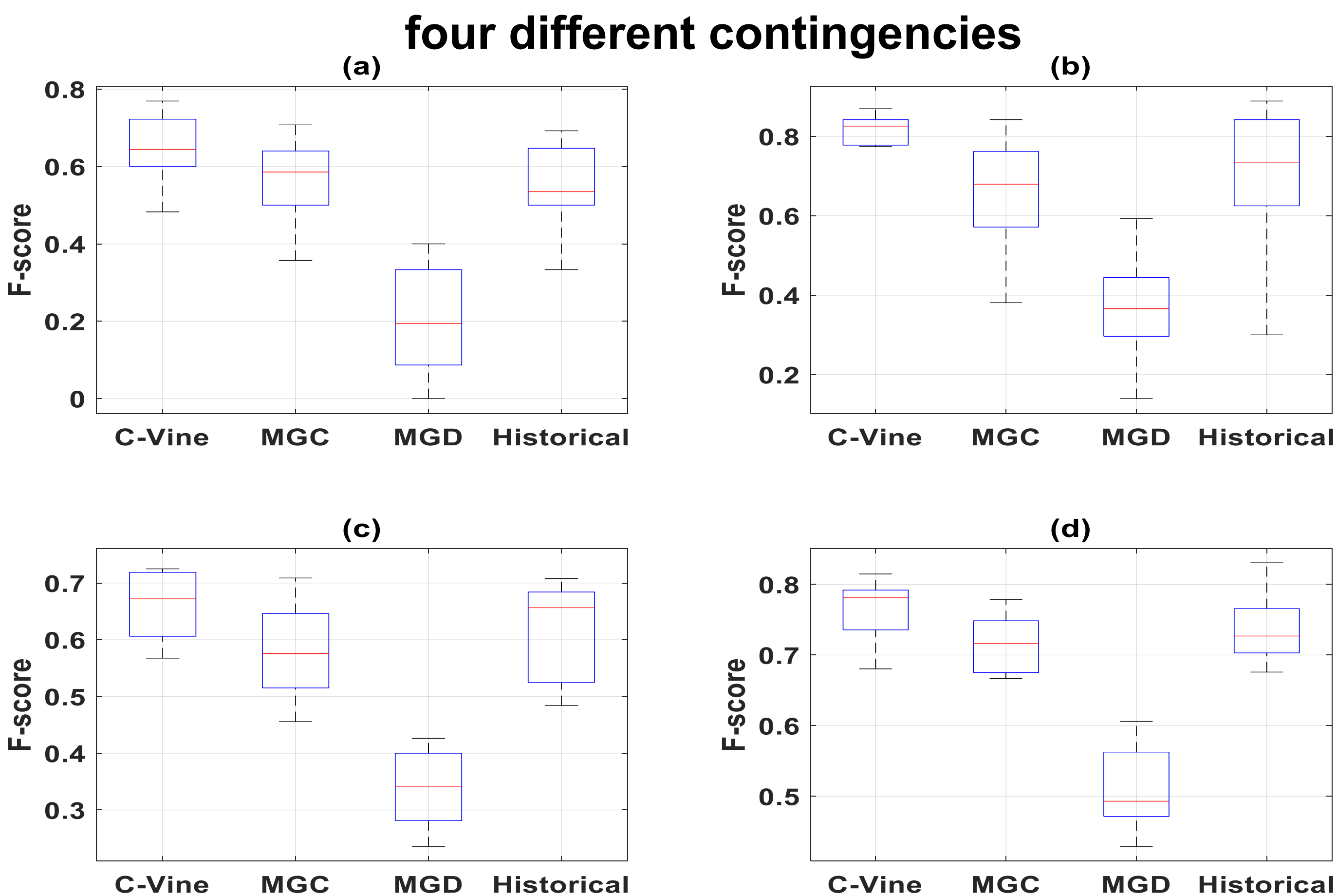
- 128 variables (118 loads, 10 wind generators)
- 3 months data, 15-minute intervals
- 10 clusters; 97.5% variance used to select truncation; C-vine parametrised using Clayton, Frank, Gaussian, Gumbel, Student-t copulas (and rotations)

- Generated 40,000 samples
- Test on 1000 random subsets of sample and historical data
- Test metrics
  - Kolmogorov-Smirnov for marginals
  - Energy test (Aslan & Zech, 2005)

Statistical match to historical data



Impact on machine learning



MGC = multivariate Gaussian copula ; MGD = multivariate Gaussian distribution

Scaling up

- Realistic system (French grid)
  - 1886 buses, 1955 lines
  - 3808 variables
  - 1980 credible contingencies
- HPC implementation (10,000 cores; 223,500 core-hours)
  - 9870 random samples processed
  - 14M dynamic simulations
  - 1.35 TB of impact analysis data (for machine learning)

Konstantelos et al., *IEEE Trans Smart Grid*, 8(3), pp.1417–1426 (2017).

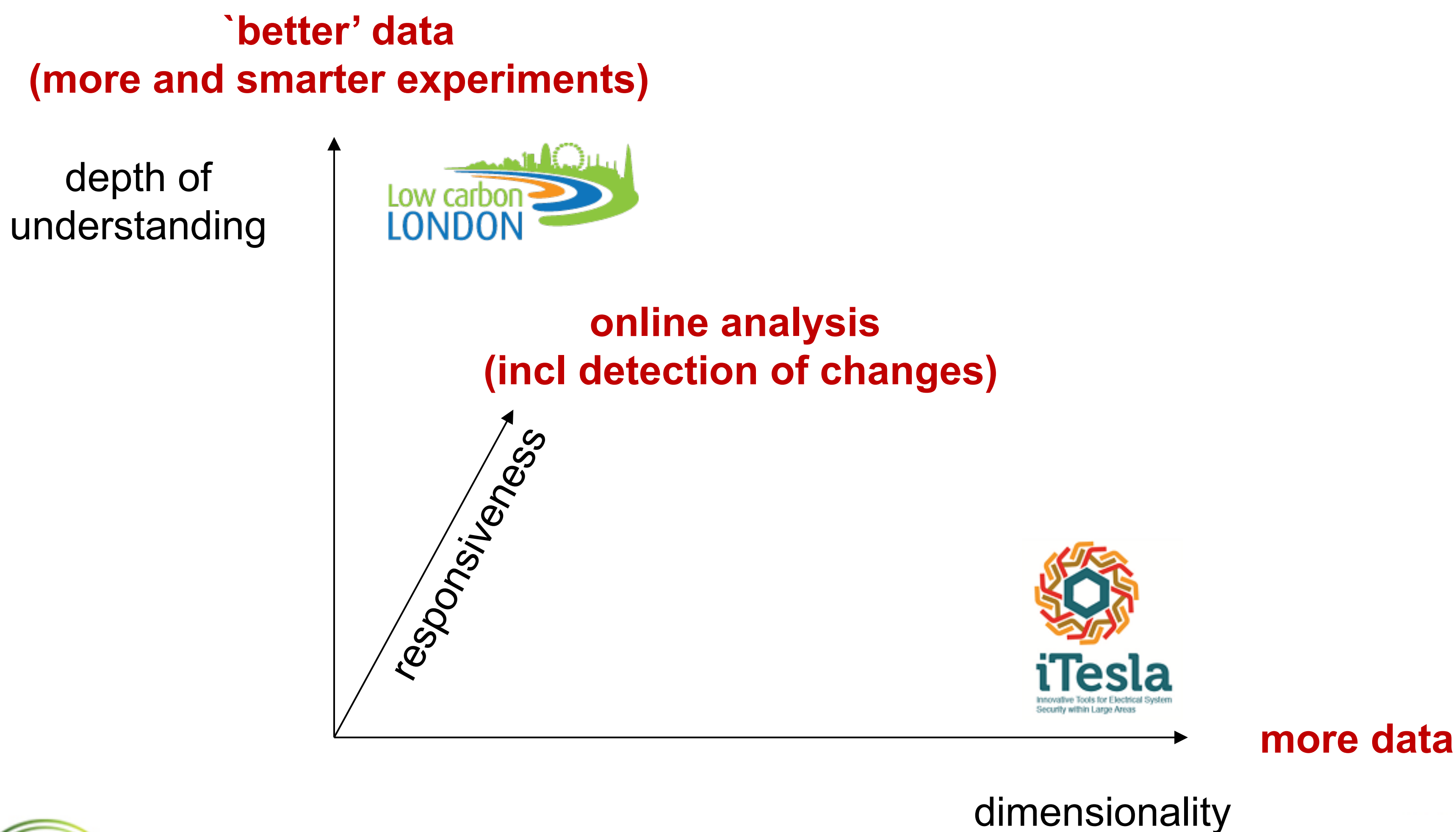
Summary

- Smart meter data
  - Quantify aggregate demand response - and uncertainty
  - Quantify household responsiveness probabilistically

- Transmission level data
  - Data driven model to sample ‘typical’ scenarios

- Data volume may not be Big, but is large and growing

Challenges in load modelling



IEEE PES Power & Energy Society®

IEEE

Funders

This research was supported by the iTesla project within the 7th European Community Framework Programme

SEVENTH FRAMEWORK PROGRAMME

"Low Carbon London" was funded through the Low Carbon Networks Fund programme, administered by the UK Regulator, Ofgem.

ofgem

Low carbon LONDON

IEEE PES Power & Energy Society®

IEEE

Thank you

- **Low Carbon London Project: Data from the Dynamic Time-of-Use Electricity Pricing Trial, 2013**
  James Schofield, Richard Carmichael, Simon Tindemans, Mark Bilton, Matt Woolf, Goran Strbac. (2016). UK Data Service. SN: 7857
- **A baseline-free method to identify responsive customers on dynamic time-of-use tariffs**
  James Schofield, Simon Tindemans, Goran Strbac
  arXiv:1605.08078
- **Resilience performance of smart distribution networks**
  Simon Tindemans, Predrag Djapic, James Schofield, Tatiana Ustinova and Goran Strbac
  Report D4 for the "Low Carbon London" LCNF project, 2014.
- **Residential consumer responsiveness to time varying pricing**
  James Schofield, Richard Carmichael, Simon Tindemans, Matt Woolf, Mark Bilton and Goran Strbac
  Report A3 for the "Low Carbon London" LCNF project, 2014.
- **Implementation of a Massively Parallel Dynamic Security Assessment Platform for Large-Scale Grids**
  Ioannis Konstantelos et al., IEEE Transactions on Smart Grid 8(3), 2017.
- **Evaluating composite approaches to modelling high-dimensional stochastic variables in power systems**
  Mingyang Sun, Ioannis Konstantelos, Simon Tindemans and Goran Strbac
  PSCC 2016
- **C-Vine Copula Mixture Model for Clustering of Residential Electrical Load Pattern Data**
  Mingyang Sun, Ioannis Konstantelos and Goran Strbac
  IEEE Transactions on Power Systems 32(3), 2017.