Data Analytics

Course: 18-787

Patrick McSharry
patrick@mcsharry.net
www.mcsharry.net
Twitter: @patrickmcsharry

Spring 2023

ICT Center of Excellence
Carnegie Mellon University

Data Analytics

WEEK 3A

Course outline

Week	Lecture A	Lecture B
1	Data Analytics	Weather forecasting
2	Renewable energy	Wind energy
3	Solar energy	Demand forecasting
4	Risk	Extreme events
5	Health	Biomedicine
6	Early warning systems	Economic forecasting

A2

- Q3: wind power fluctuation graph (Week2b)
- Q4: show quantiles of $r(t,d)$ for each time scale d
- Q7: variance ratio test: vratiotest.m
- Q8/9: generate forecasts $x(t+1)^* = SMA(x,k)$. Note that persistence is $SMA(x,k)$ with $k=1$. Compare forecast performance (MAE) versus k.
- Q10: consider ARIMA(p,d,q) and find optimal values p,d,q to minimize MAE for one-step forecast.

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Storage and price	10
2	Discussion	PV and thermal	10
3	Case study	Actual solar efficiency	10
4	Analysis	Efficiency versus irradiance	20
5	Demo	Smoothing techniques	20
6	Q&A	Questions and feedback	10

Solar Energy

- Solar thermal
  - Solar thermal electric energy generation concentrates the light from the sun to create heat, and that heat is used to run a heat engine, which turns a generator to make electricity
- Solar photovoltaic (PV)
  - PV directly converts the sun's light into electricity and is only effective during daylight hours because storing electricity is not a particularly efficient process

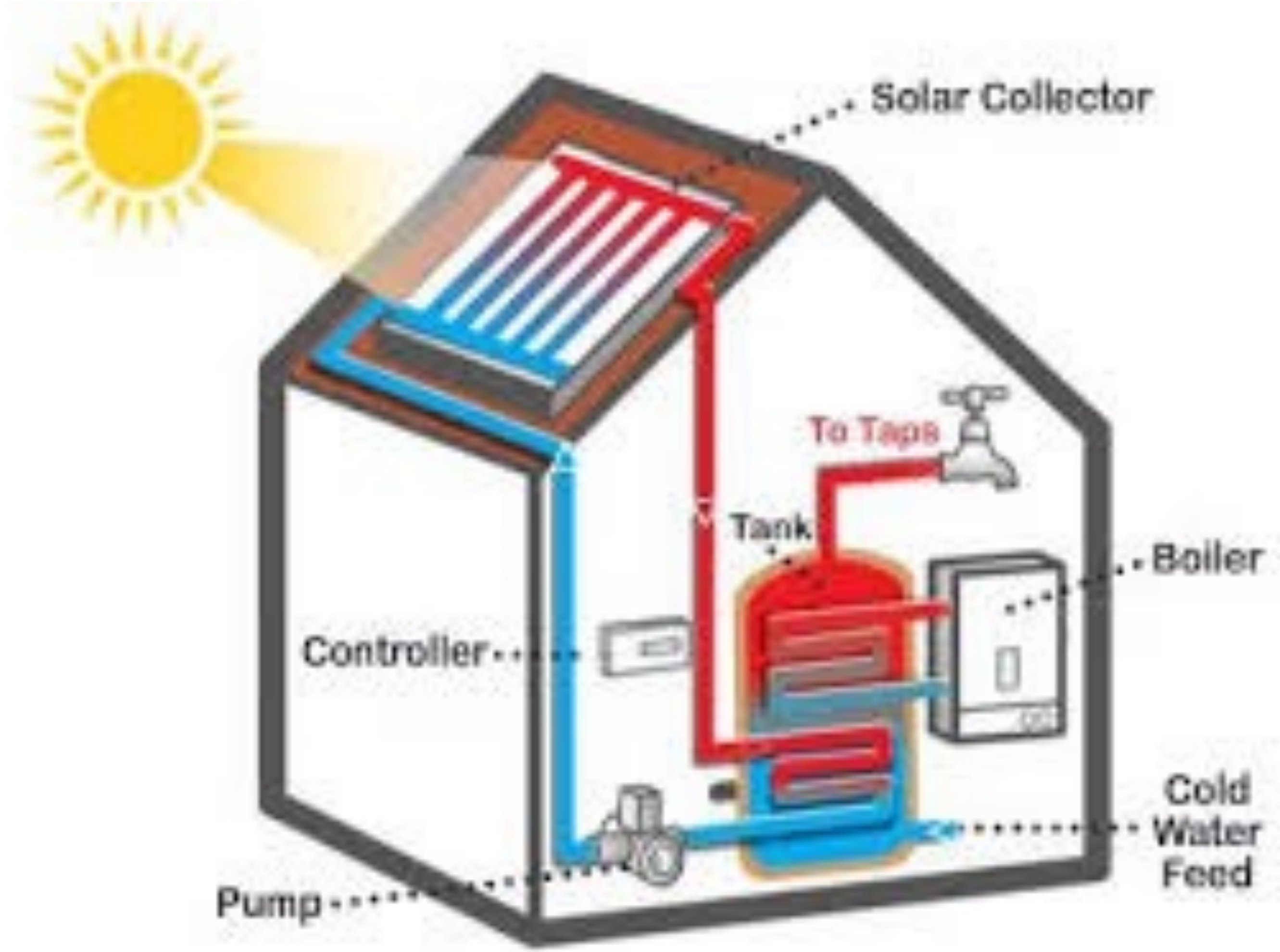
Efficiency considerations

- Heat storage is easier and more efficient than storing electricity.
- Therefore <u>solar thermal is much more attractive for large-scale energy production</u>.
- <u>Heat can be stored during the day and then converted into electricity at night</u>.
- Solar thermal plants with storage capacity can greatly improve its economic feasibility.

Solar thermal

Source: www.solarpowerportal.co.uk

Solar thermal schematic



Source: www.lifesenergy.co.uk

Ivanpah, California (392 MW)

Source: www.technocrazed.com

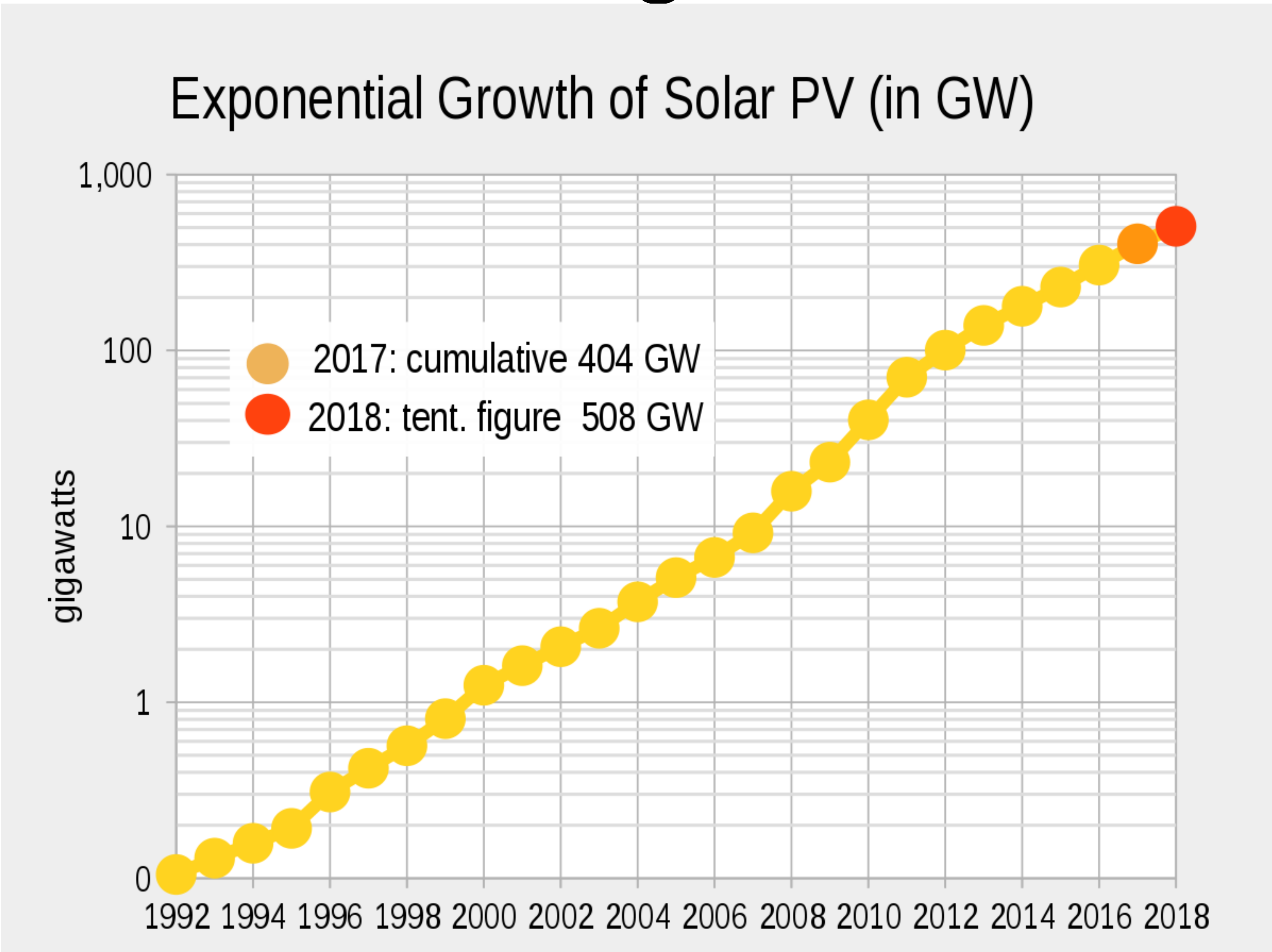
Solanova, Spain (150 MW)

Source: www.wikipedia.com

Agahozo Shalom Youth Village, Rwanda

8.5 MW solar PV (6% Rwanda's capacity); $23.7 million project of 17 hectares is the first utility-scale, grid-connected, commercial solar field in East Africa.

Global growth

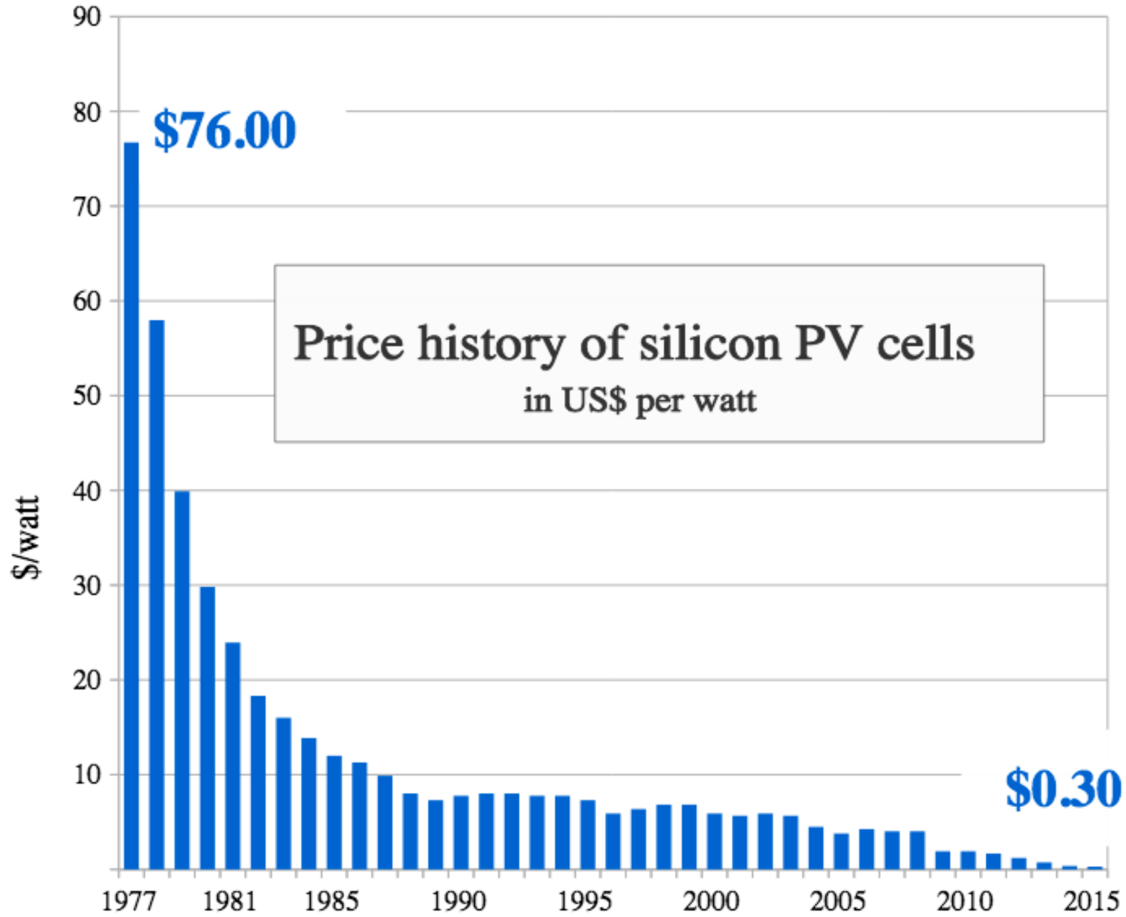


Moore’s Law

- Moore's law is the observation that the number of transistors in a dense integrated circuit (IC) doubles about every n years, where n is:
  a) One
  b) Two
  c) Three
  d) Four

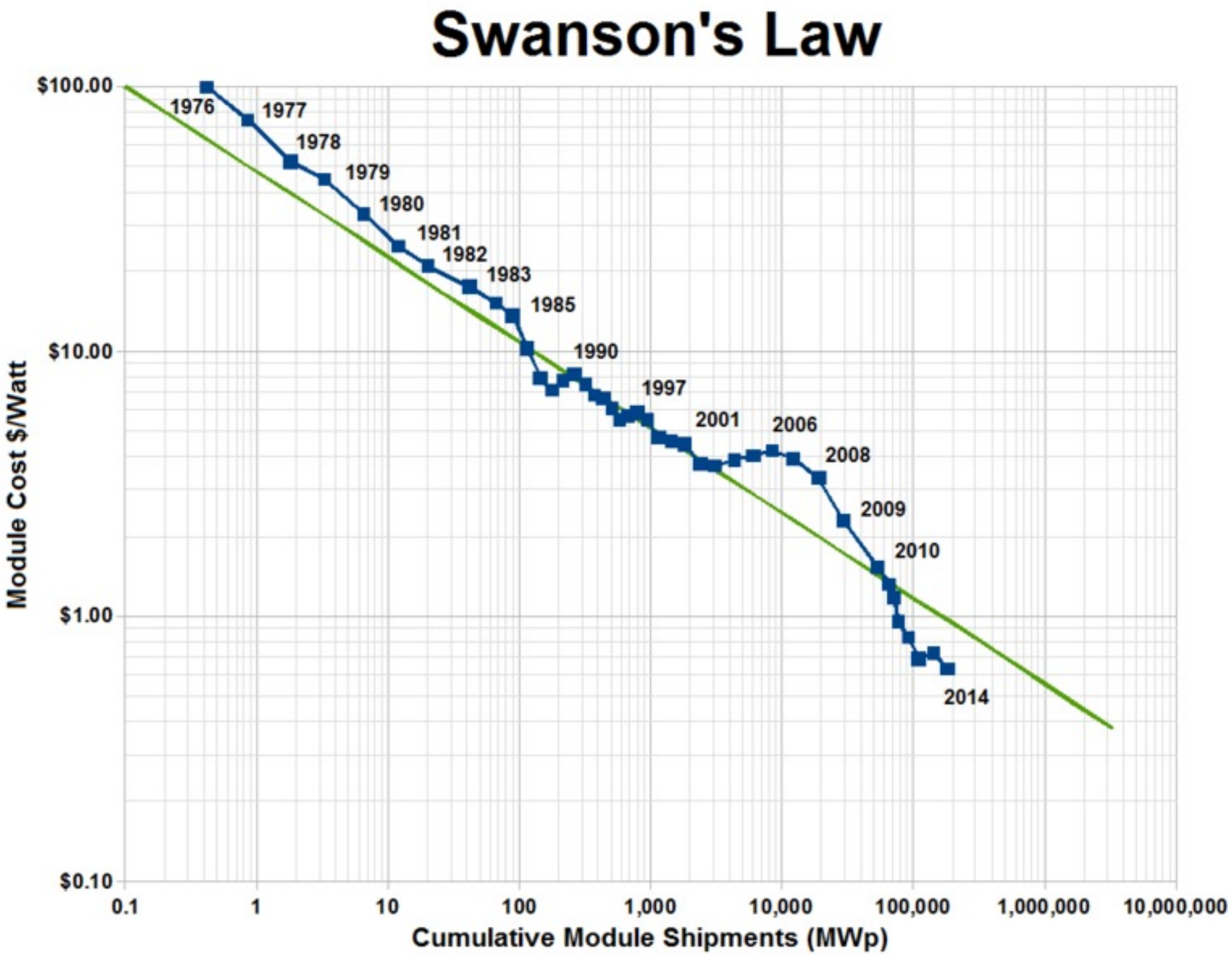
- **Slido.com #63347**

Price of Photovoltaics



Source: Bloomberg New Energy Finance & pv.energytrend.com

Swanson’s Law (aka Wright’s Law)



Price of solar photovoltaic modules tends to drop 20% for every doubling of cumulative shipped volume (Richard Swanson, founder of SunPower Corporation).

Wright’s law is more appropriate than Moore’s law for policymaking since costs can be driven down by boosting cumulative production (Bailey et al., 2017)

A.G. Bailey, J.D. Bakker, D. Rebois, R. Zadourian, P. McSharry, JD Farmer. (2017) How well do experience curves predict technological progress? A method for making distributional forecasts. Technological Forecasting and Social Change.

Community Photovoltaics

- **Oxford North Community Renewables (ONCORE)** is an Industrial and Provident Society which was set up to fund the installation of solar photovoltaic (PV)
- Raised £160,000 for 47 KW installation
- Objectives:
  - generate clean renewable solar power
  - make our schools greener and give them cheaper electricity
  - provide funds for low carbon projects
  - give investors 4-7% returns

Source: www.oncore.org.uk

Solar Photovoltaics

Source: www.oncore.org.uk

Solar Irradiance

- Solar power is produced by the Sun in the form of electromagnetic radiation, which is perceived by humans as sunlight.
- Solar irradiance is a measure of the solar radiation **power** per unit area on the Earth's surface (Wm^{-2}).
- Solar insolation is the total amount of solar radiation **energy** received on a given surface area during a given time (Jm^{-2}).

SI Units

- Energy: Joule (J)
- Power: Watt (W)
- Electricity consumption (electrical energy) is also measured as the amount of power used over a specific time period.
- Using hours, we have 1 kWh = 3.6 MJ
- A 1 kW heating device consumes 1 kWh of electrical energy in 1 hour.
- Irradiance: power per metre squared (Wm^{-2})
- Insolation: energy per metre squared (Jm^{-2},$kWhm^{-2}$)

Solar PV Tariffs Worldwide

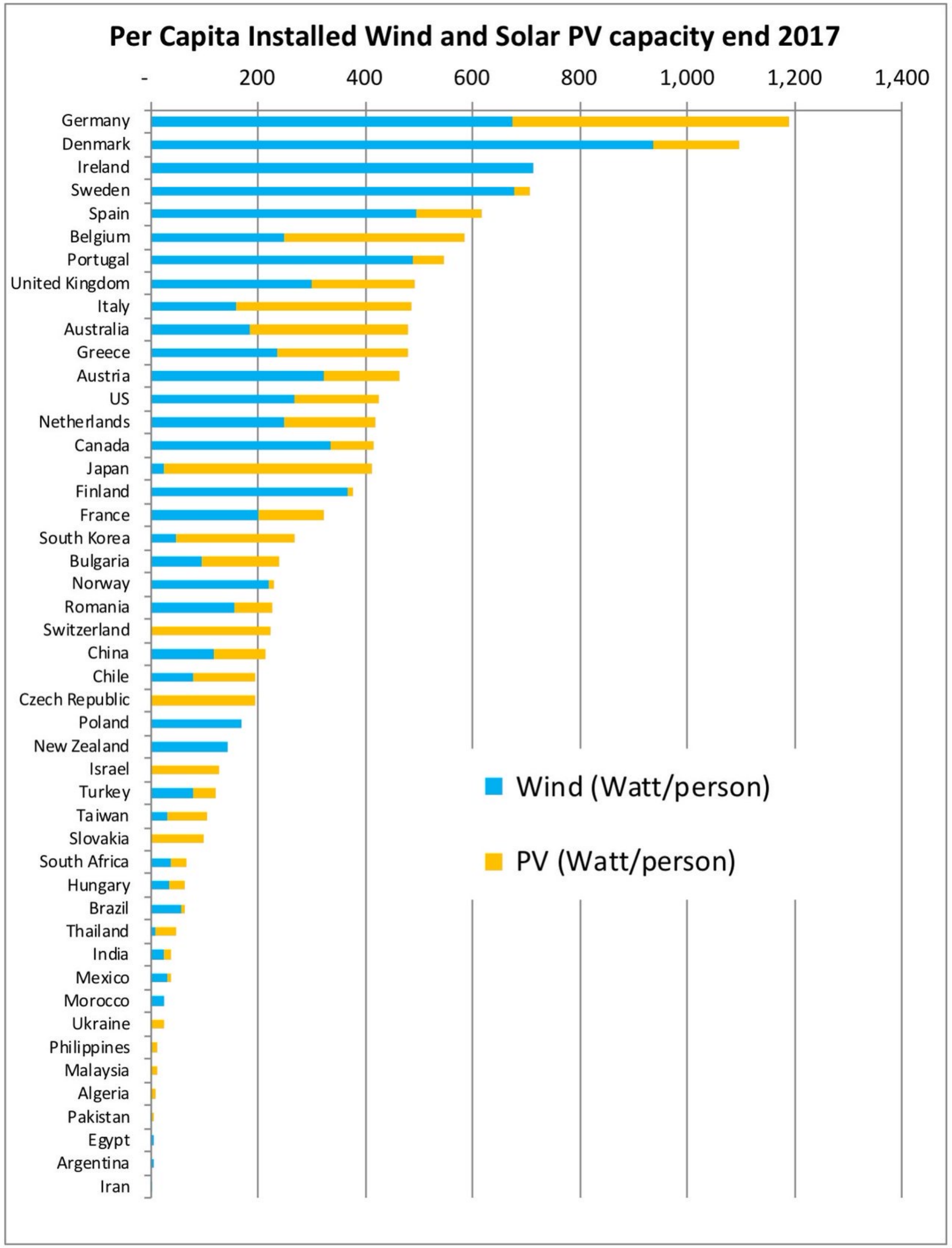
Price Summary for Solar PV Tariffs Worldwide				
20-Sep-11			1.387	1.350
Selected programs with contract terms of 15 years or longer.				
Jurisdiction	Years	€/kWh	CAD/kWh	USD/kWh
Small or Non-Differentiated				
Switzterland	25	0.590	0.819	0.796
Ontario MicroFIT	20	0.578	0.802	0.780
Great Britain	20	0.425	0.584	0.606
France	20	0.406	0.644	0.626
Bulgaria	25	0.389	0.539	0.525
Luxembourg	15	0.382	0.530	0.516
Slovakia	15	0.382	0.530	0.516
Italy September 2011	20	0.361	0.501	0.487
Cyprus	20	0.360	0.499	0.486
Israel	20	0.334	0.464	0.451
Slovenia	15	0.332	0.461	0.449
Czech Republic	20	0.304	0.422	0.411
Ecuador	15	0.297	0.411	0.400
Malaysia	21	0.288	0.399	0.389
Germany	20	0.287	0.399	0.388
Uganda	20	0.268	0.372	0.362
Rhode Island	15	0.247	0.343	0.334
Gainesville, FL	20	0.237	0.329	0.320
Vermont	25	0.222	0.308	0.300
Hawaii All Islands	20	0.162	0.224	0.218
Large				
Switzterland	25	0.353	0.490	0.477
Ontario	20	0.319	0.443	0.431
Israel	20	0.298	0.414	0.402
Slovenia	15	0.288	0.399	0.388
Italy September 2011	20	0.264	0.366	0.356
Germany	20	0.221	0.306	0.298
Malaysia	21	0.199	0.276	0.269
France <12 MW	20	0.120	0.166	0.162

Source: www.wind-works.org

Poll: Solar PV installation

- Which country had the highest **per capita** solar PV installed in 2017?
  a) China
  b) Germany
  c) Japan
  d) USA

Slido.com #63347



Note the different installation levels by solar PV and wind.

The countries that are leading on solar PV do not necessarily have the greatest solar resources.

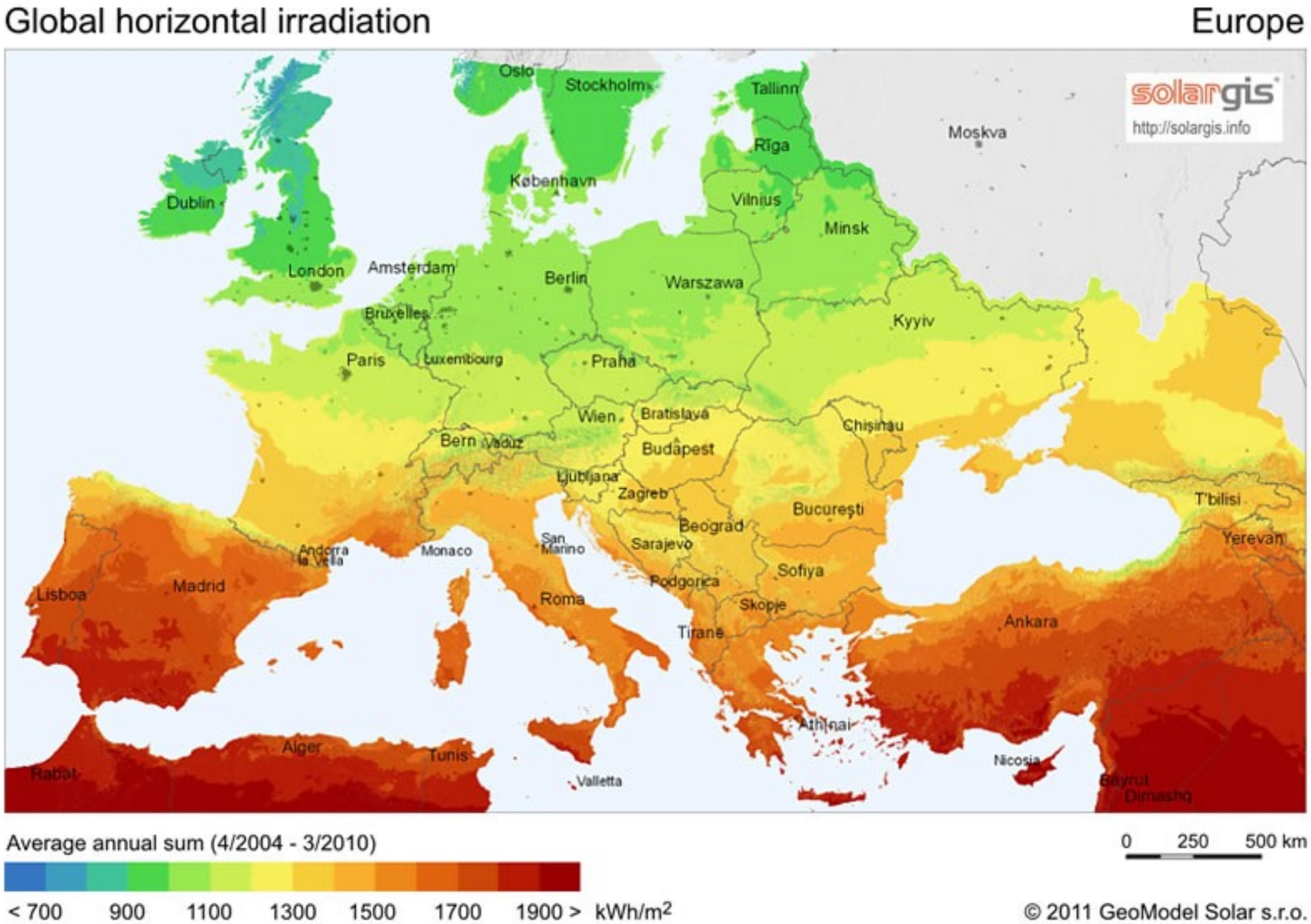
Source: BP Statistical Review of World Energy 2018

@RobertKleiburg

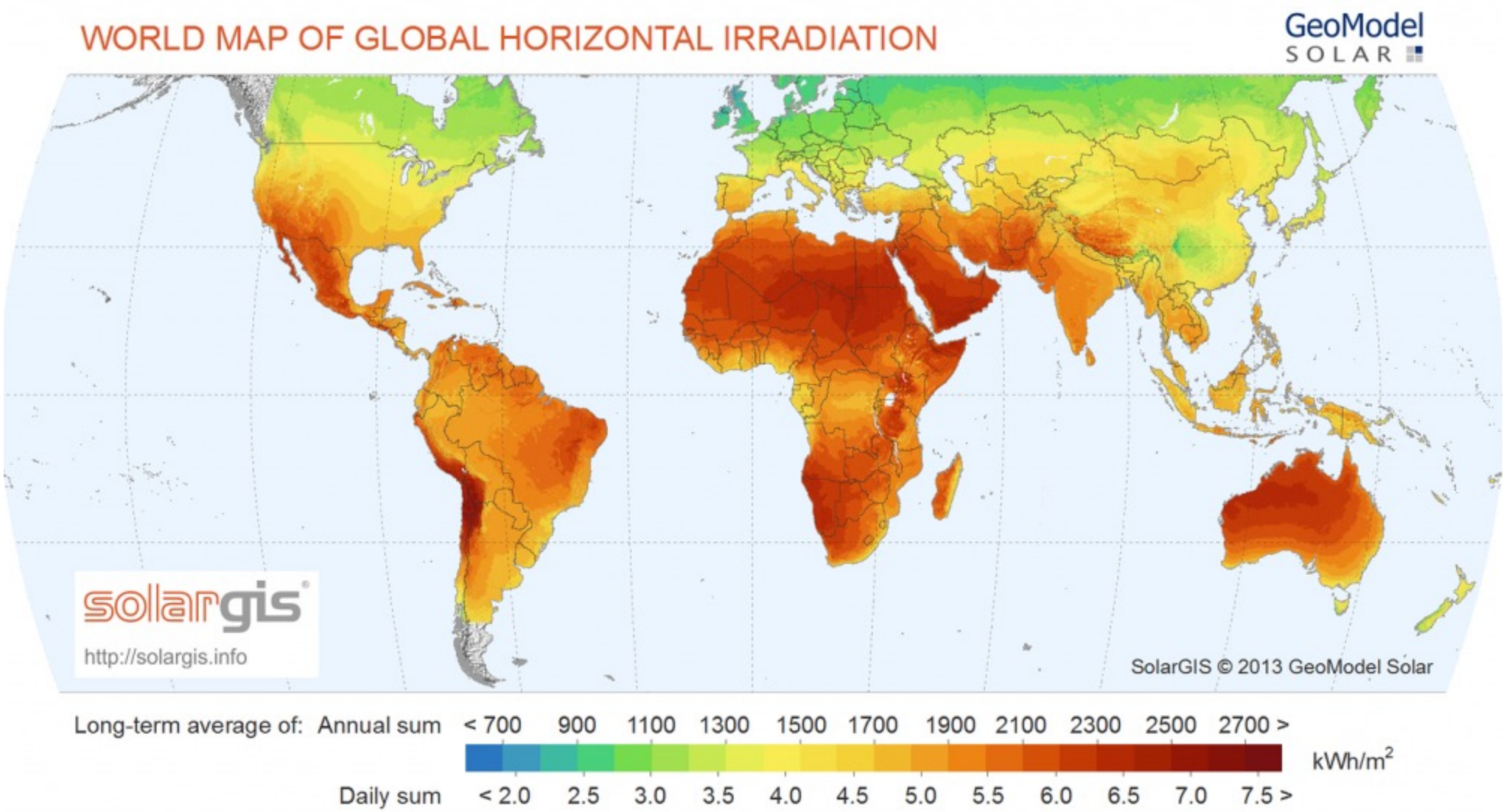
Solar constant

- Over the course of a year the average solar radiation arriving at the top of the Earth's atmosphere at any point in time is roughly 1366 Wm^{-2} and is known as the solar constant.
- The Sun's rays are attenuated as they pass through the atmosphere, thus reducing the irradiance at the Earth's surface to approximately 1000 Wm^{-2} for a surface perpendicular to the Sun's rays at sea level on a clear day.

Europe Irradiation Map



World Irradiation Map



Morocco

- Solar power in Morocco is enabled by the country having one of the highest rates of solar insolation among other countries— about 3,000 hours per year of sunshine but up to 3,600 hours in the desert.
- Morocco has launched one of the world's largest solar energy projects costing an estimated $9 billion, aiming to create 2,000 megawatts of solar generation capacity using both photovoltaic and concentrated solar power technology.
- Morocco, the only African country to have a power cable link to Europe (2,100 MW), aims to benefit from the €400bn ($573.8bn) expected to come from the ambitious pan-continental Desertec Industrial Initiative.

Solar energy output

- Actual output depends on factors that depend on the geographical location and specific site.
- Factors include:
  - Intensity of the sunlight
  - Number of daylight hours
  - Percentage of hours where clouds prevent direct sunlight
  - Orientation and tilt of the installation
  - Local effects that cause shadows (e.g. trees or buildings)

PV-Compare

- Environmental Change Unit, Oxford University project to compare eleven different PV technologies in realistic conditions in the UK (Oxford) and Spain (Mallorca)
- Half-hourly recordings of meteorological conditions and the power generated
- Manufacturers provide specifications such as the peak power rating, which describes the power output under “Standard Test Conditions” (STC).
- These conditions are defined as an irradiance level of 1,000 Wm^{-2}, an air mass spectral distribution of 1.5 and a cell temperature of 25 ^{o}C

STC versus reality

- In practice, however, these ideal STC are unlikely to be met. In fact an irradiance level as high as 1,000 Wm^{-2} is representative of summer-time clear sky conditions.
- Under such conditions, the cell temperature could reach 50 ^{o}C, thereby reducing the performance from that quoted under the STC.
- The effects of combinations of temperature, cloud cover, elevation of the sun and other climatic conditions are not known. This lack of information complicates the process of choosing a photovoltaic technology to suit specific climatic conditions.

PV-Compare Data

- Site locations:
  - Begbrook, Oxford, UK (1W, 52N)
  - S'Alqueria, Mallorca, Spain (3E, 39N).
- Variables measured every half hour:
  - Power output of eleven PV technologies,
  - Irradiance (insolation),
  - Three directional insolation measurements,
  - Ambient temperature,
  - Temperature of the modules.

Word Cloud: Missing data

- What techniques would you recommend for dealing with missing data?

Slido.com #63347

Missing Data

- We used a variety of approaches for filling gaps in the data sets:
  - Univariate linear interpolation
  - Multivariate regression
- The selected approach depended on the size of the gap and the specific variables that had missing values at the particular time of the gap

PV Commercial Products

Product	Technology	Peak power, W	Area, m^2	STC efficiency	Field efficiency
Unisolar US64	Amorphous Si (Triple Junction)	512	8.1	6.32	5-6
ASE 30 DG-UT	Amorphous Si (Double Junction)	540	10.8	5	5-6
Intersolar Gold	Amorphous Si (Single Junction)	504	11.5	4.38	2-3
Solarex Millennia	Amorphous Si (Double Junction)	516	9.8	5.26	4-5
Evergreen ES 112 AC	Multicrystalline Si (Ribbon)	560	7.6	7.36	5-6
Siemens ST40	Copper indium diselenide	560	5.5	10.18	7-9
BP Solar Apollo	Cadmium Telluride	500	7.4	6.75	3-4
Astropower APX-80	Multicrystalline Si (APEX Si film)	640	8.6	7.44	4-5
BP Solar 585	Monocrystalline Si	595	4.4	13.52	10-13
Solarex MSX 64	Multicrystalline Si	640	5.6	11.42	7-9
ASE 300 DG UT	Multicrystalline Si (Edge Fed Growth)	600	4.9	12.24	8-10

Risk management

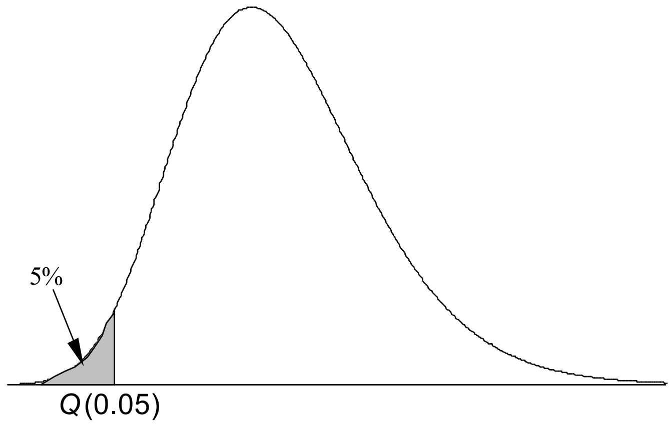
- Objective is to establish and quantify the worst case scenario
- Low levels of irradiance → no energy generated
- Typically pick an extreme quantile such as the 1% or 5%
- The 1% and 99% quantiles of the energy efficiency distribution can be used to assess and address these risks

Poll:

- Which of the following is a quantile?
  a) Mean
  b) Median
  c) Mode
  d) Variance

Slido.com #63347

Quantiles

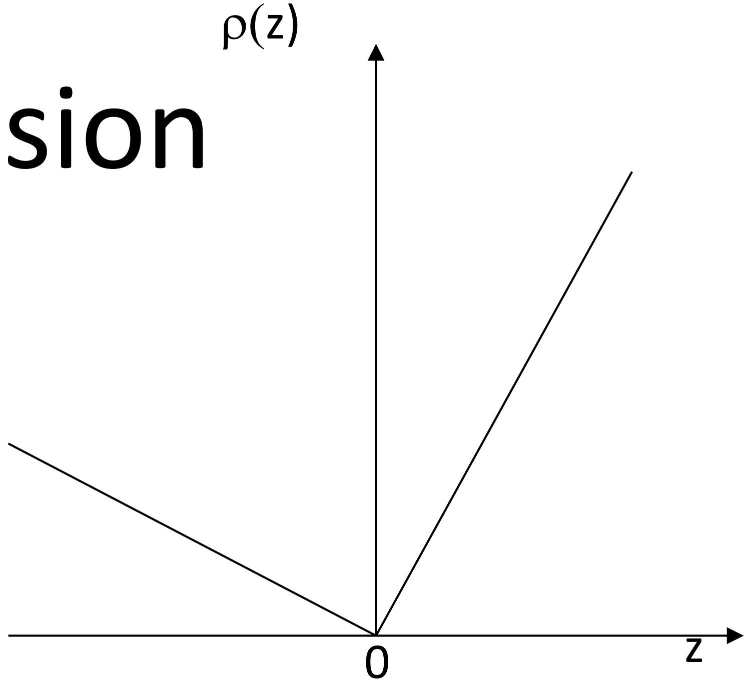


- Quantiles may be used to provide a description of the distribution

- Quantile Regression objective function:

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

Quantile regression



- Check function for quantile p:

$$\rho_p(z) = pzI_{[0,\infty)}(z) + (1-p)zI_{(-\infty,0)}(z)$$

- E.g. for estimating the median with p=0.5

Local linear quantile regression

- N observations: (X_1,Y_1), …, (X_N,Y_N). For z close to x,

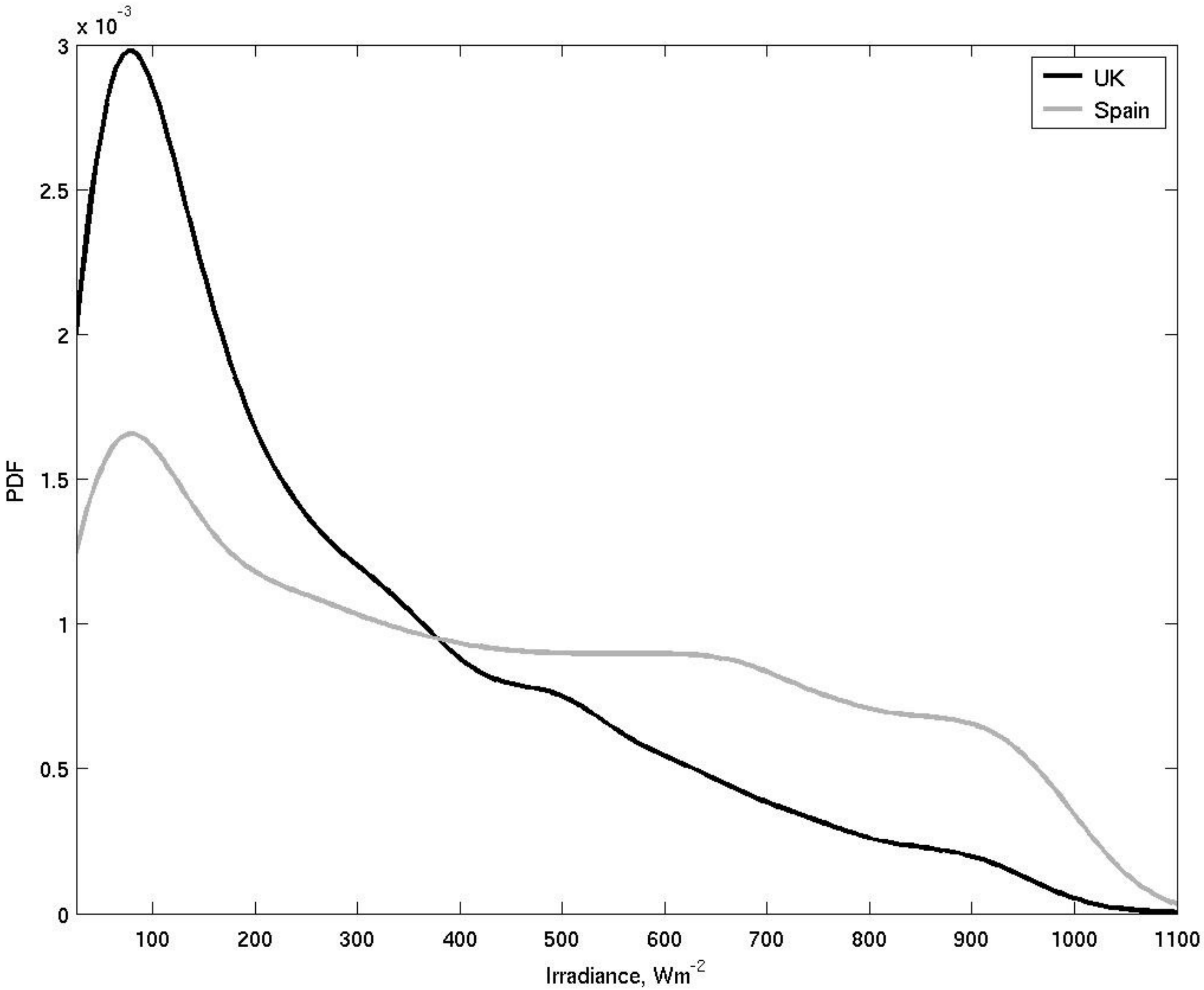
$$q_p(z) = q_p(x) + \frac{dq_p}{dx}(x)(z-x) = a + b(z-x)$$

- Estimate a and b which minimise

$$\sum_{i=1}^{n} \rho_p\big(Y_i - a - b(X_i - a)\big) K\left(\frac{x - X_i}{h}\right)$$

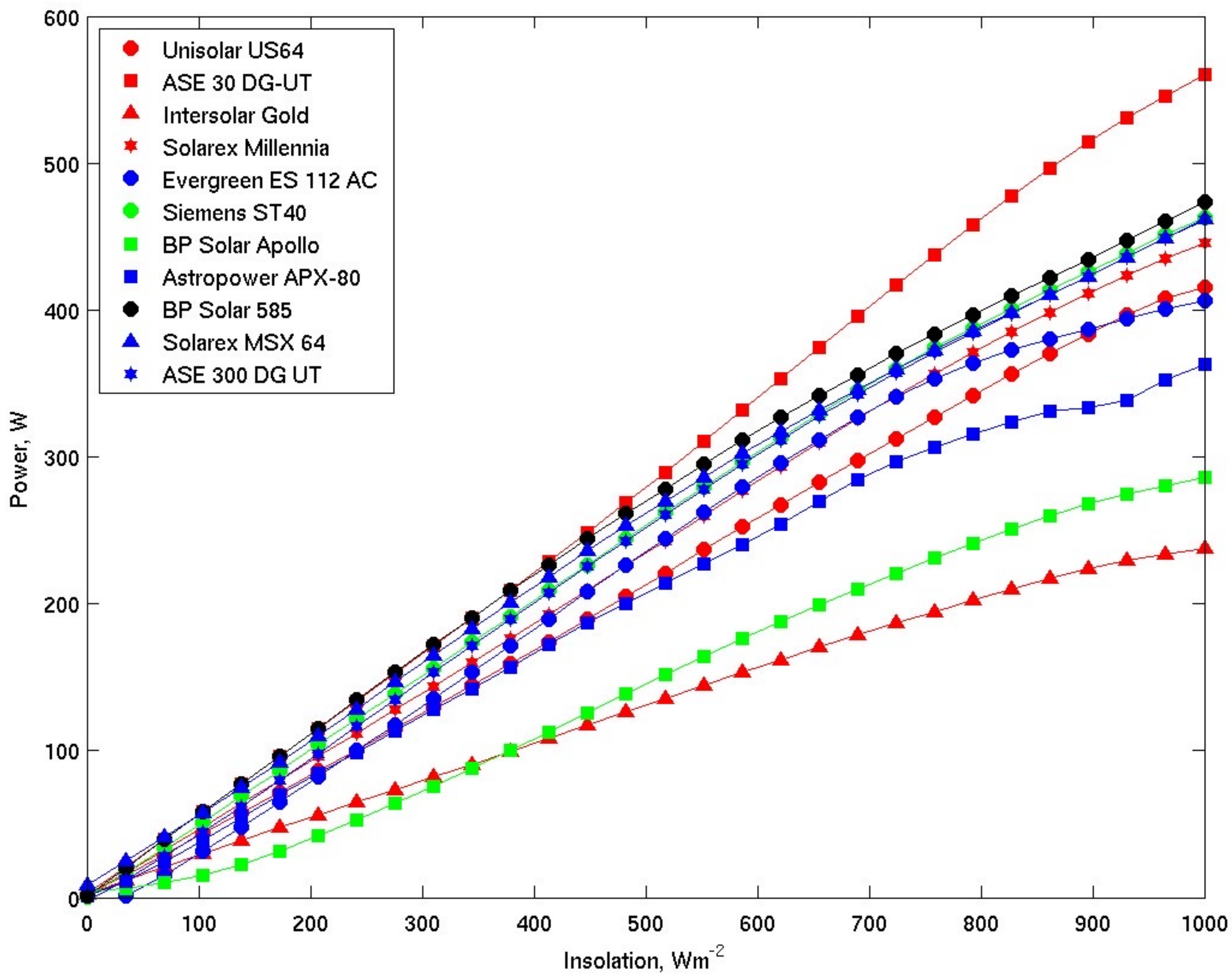
- Kernel K with bandwidth h for smoothing
- Select h using cross-validation or automatic bandwidth selection technique

Irradiance distributions

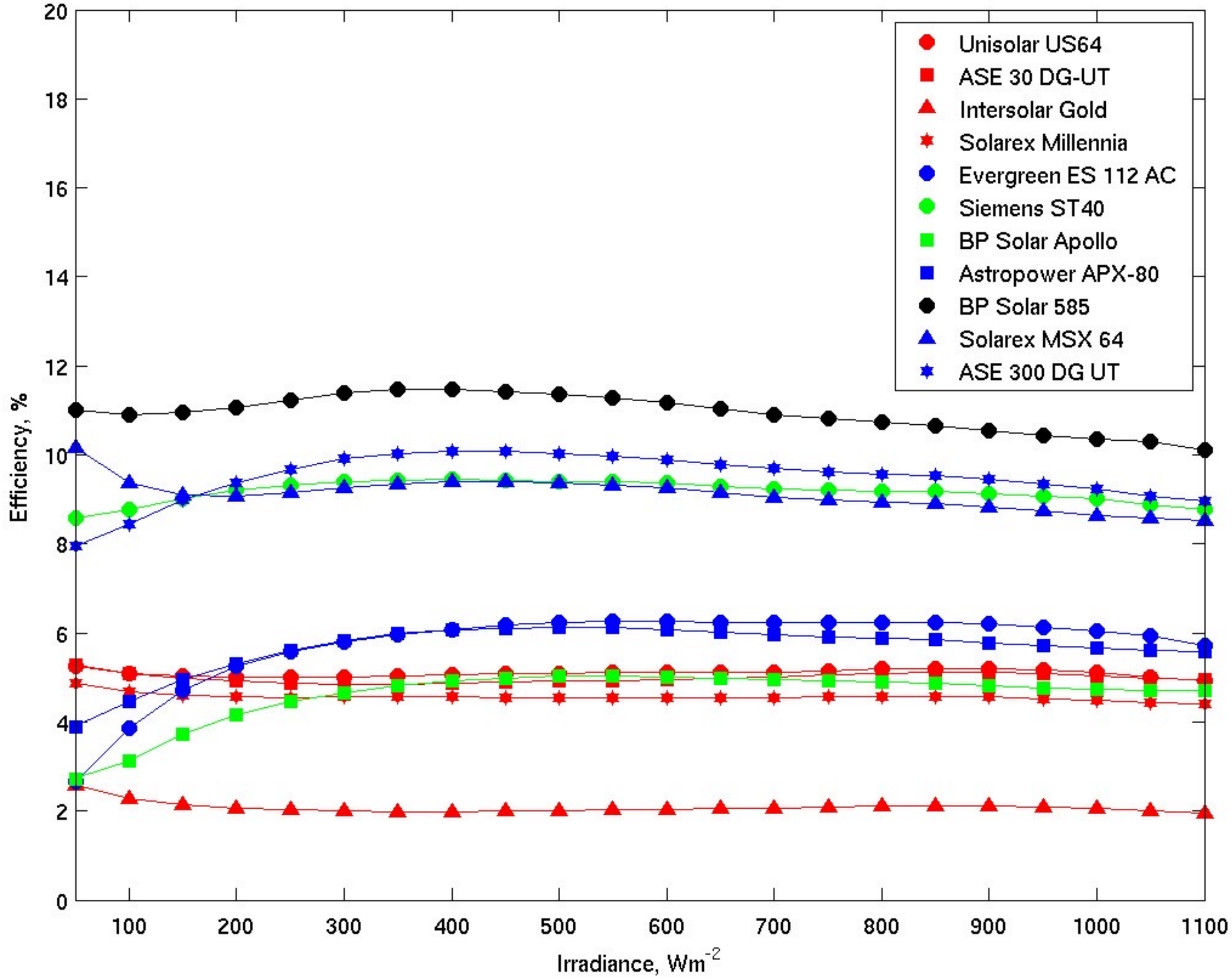


McSharry (2006). Assessing photovoltaic performance using local linear quantile regression. Proceedings of Energy and Power Systems, 165-169

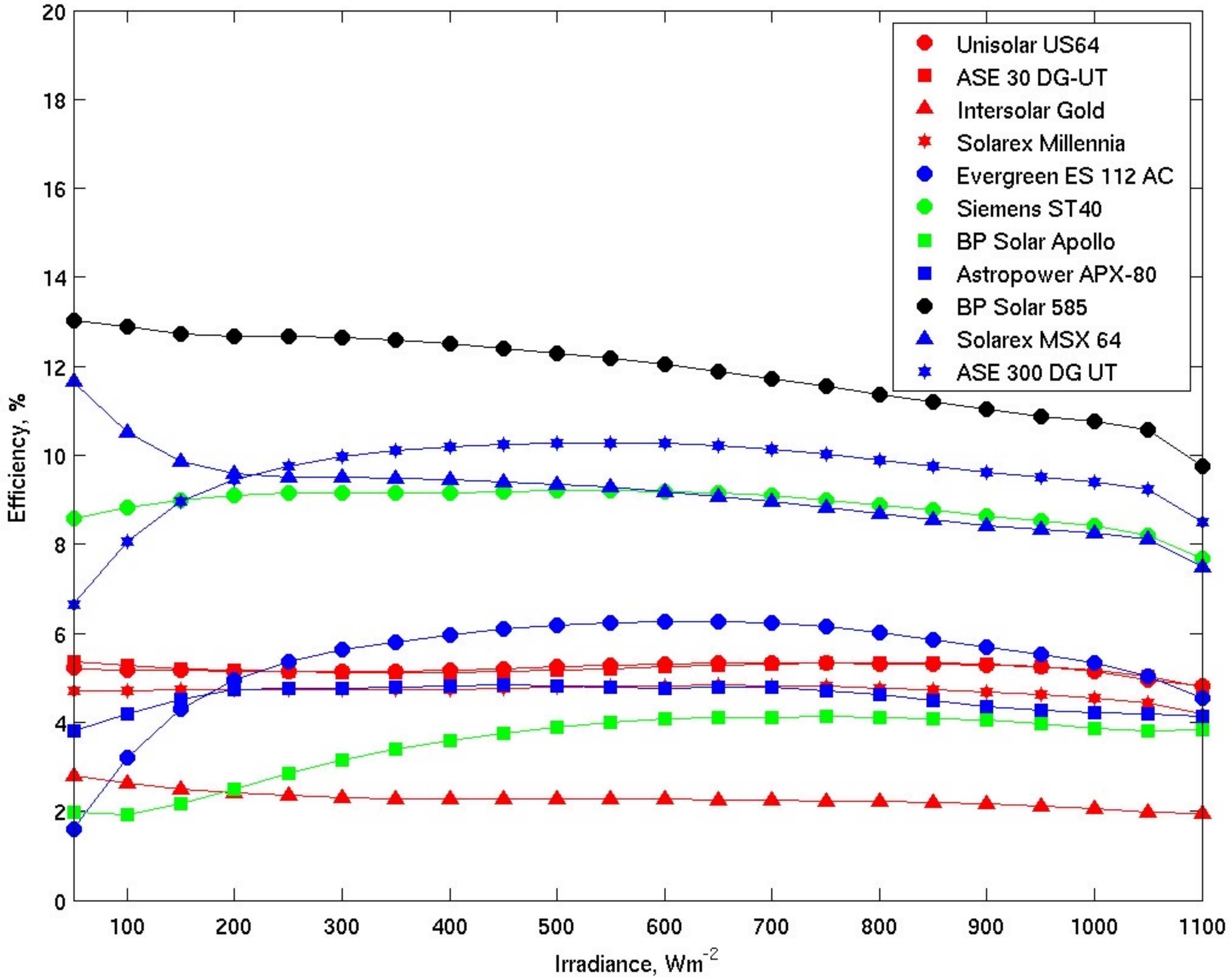
PV Power generation in Spain



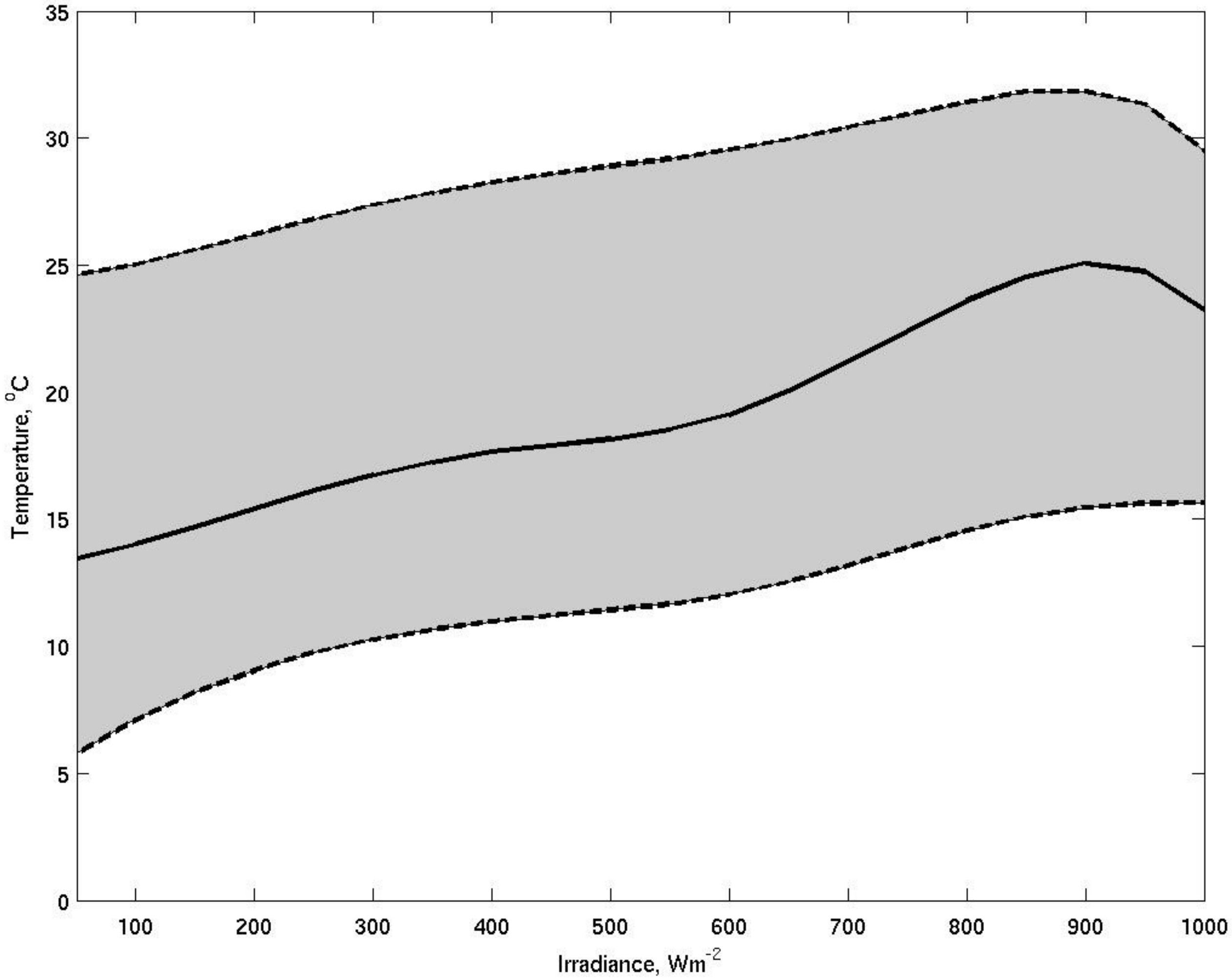
Efficiency, UK



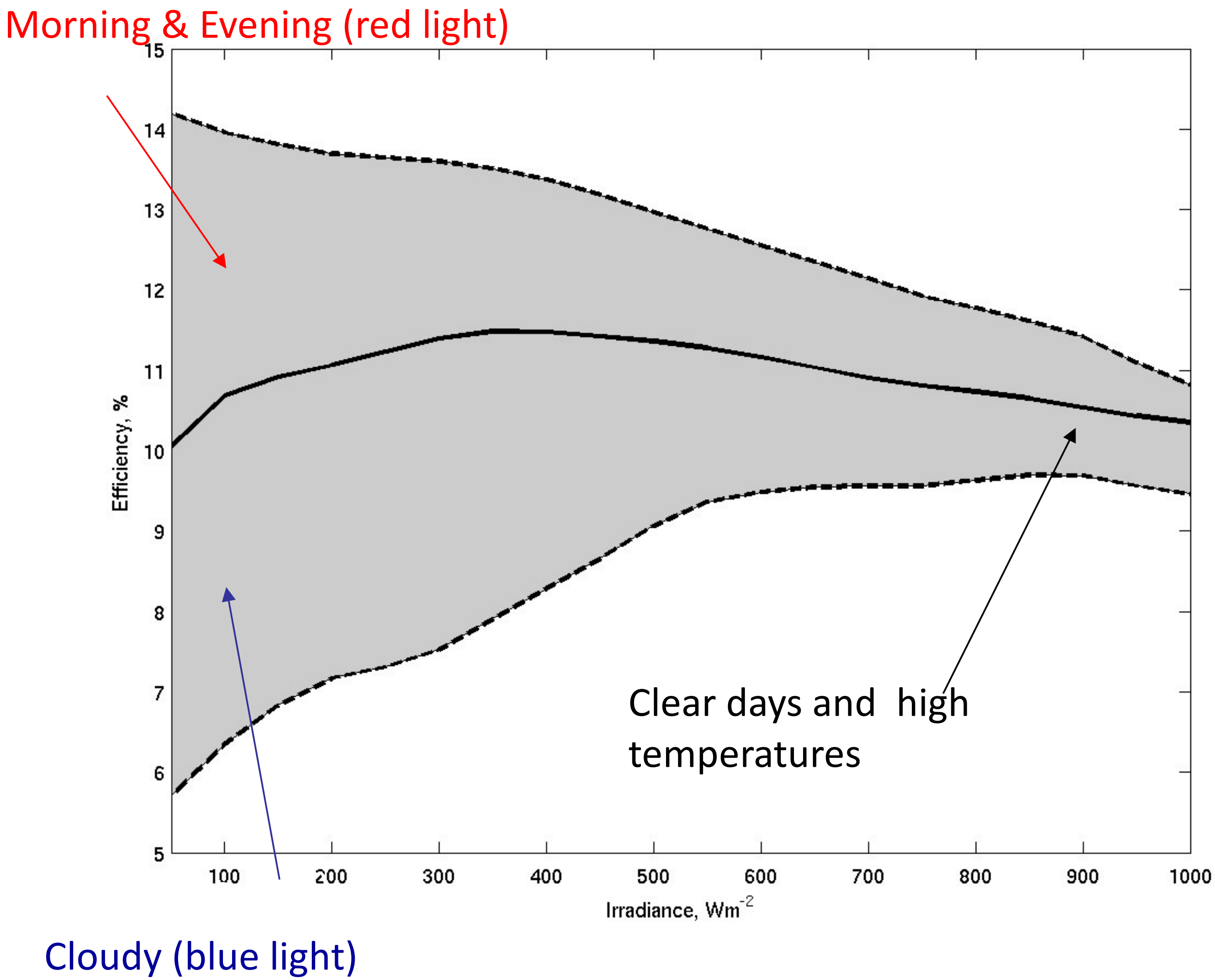
Efficiency, Spain



Temperature versus Irradiance



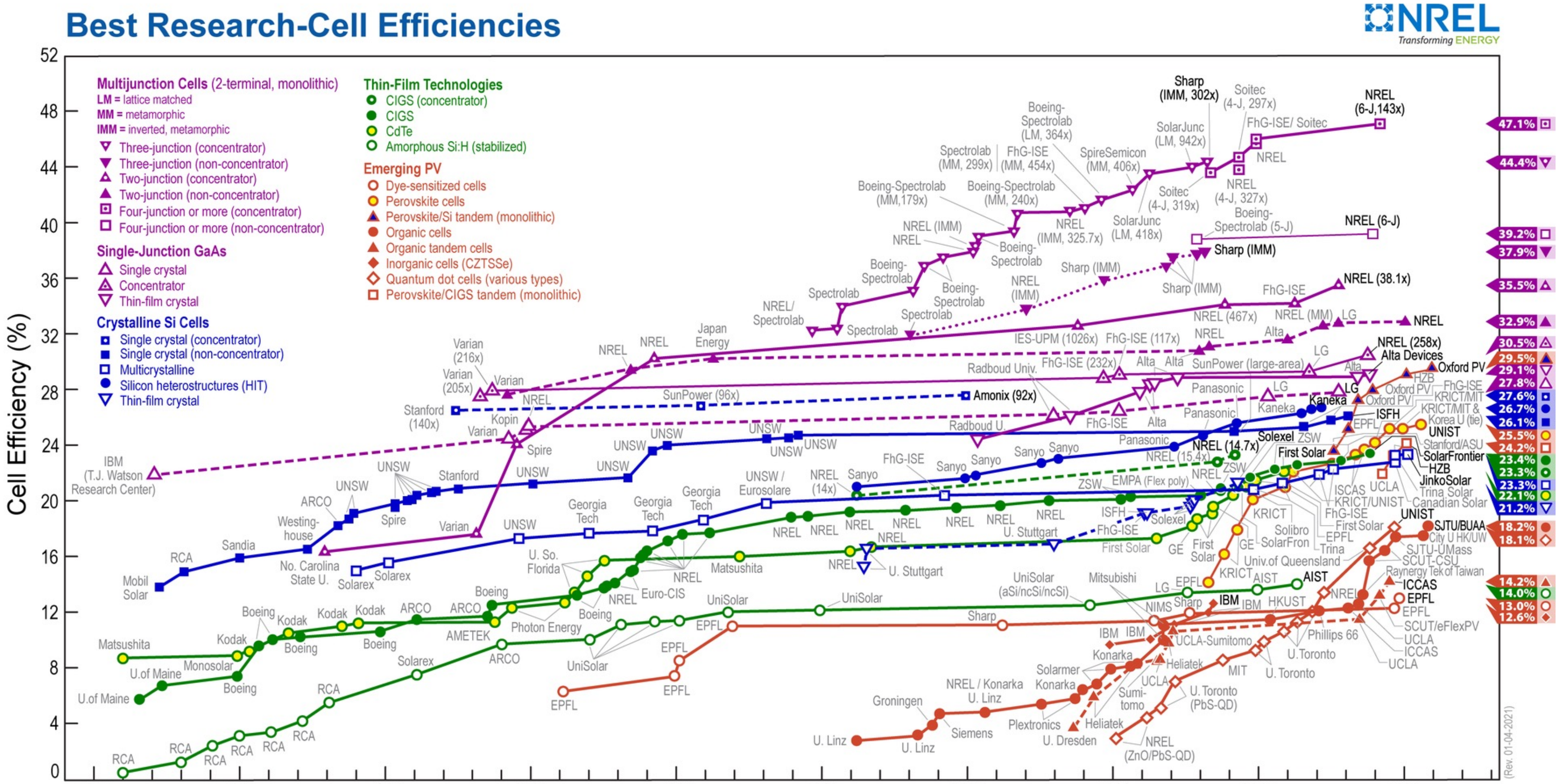
Efficiency of BP Solar 585 in the UK



Summary

- Comparison of eleven PV technologies
- Comparison across different climates
- Monocrystalline Silicon (BP Solar 585) was most efficient in both climates (efficiency of 10-13%)
- Efficiency risk assessment using LLQR
- Spectral effects explain efficiency distribution
- Extend LLQR for multiple explanatory variables

Efficiencies over time



Data Analytics

WEEK 3B

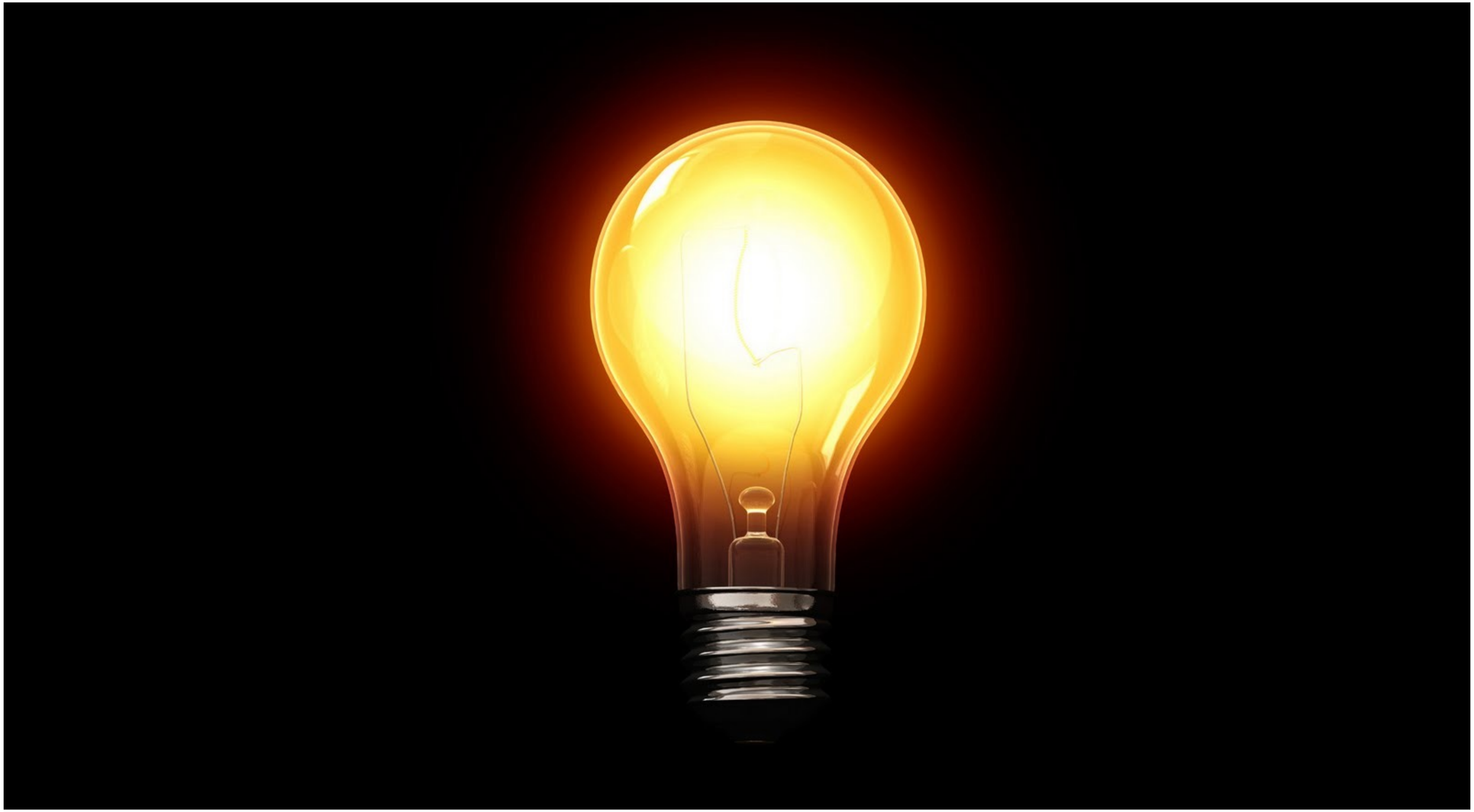
Course outline

Week	Lecture A	Lecture B
1	Data Analytics	Weather forecasting
2	Renewable energy	Wind energy
3	Solar energy	Demand forecasting
4	Risk	Extreme events
5	Health	Biomedicine
6	Early warning systems	Economic forecasting

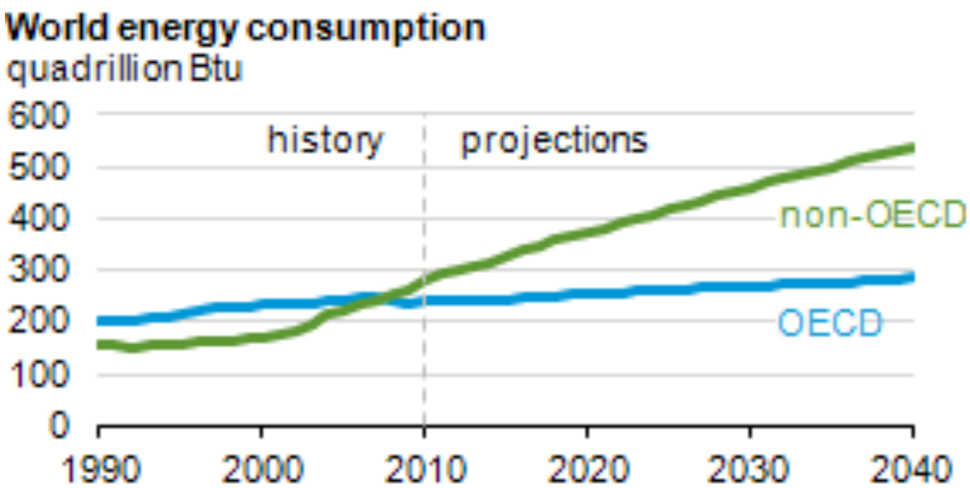
Today’s Lecture

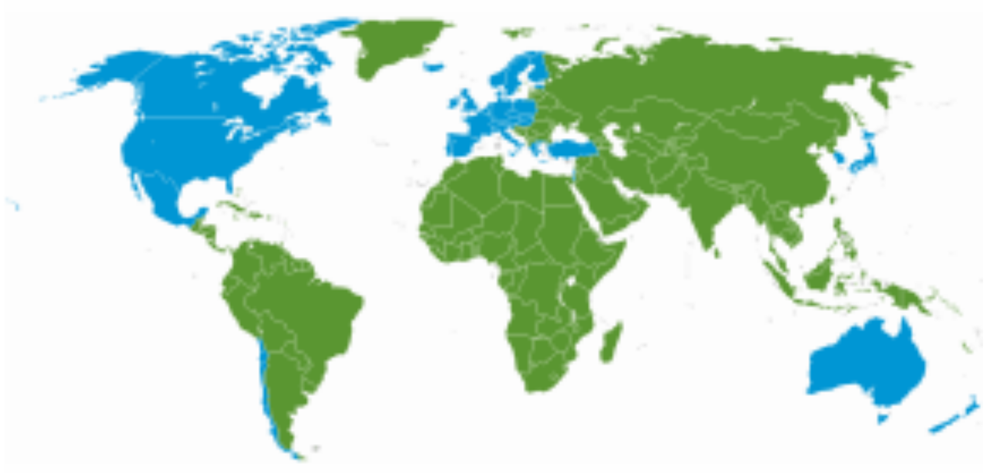
No.	Activity	Description	Time
1	Challenge	Seasonality	10
2	Discussion	Demand and weather	10
3	Case study	Peak scenarios	10
4	Analysis	Short-term forecasts	20
5	Demo	Electricity consumption	20
6	Q&A	Questions and feedback	10

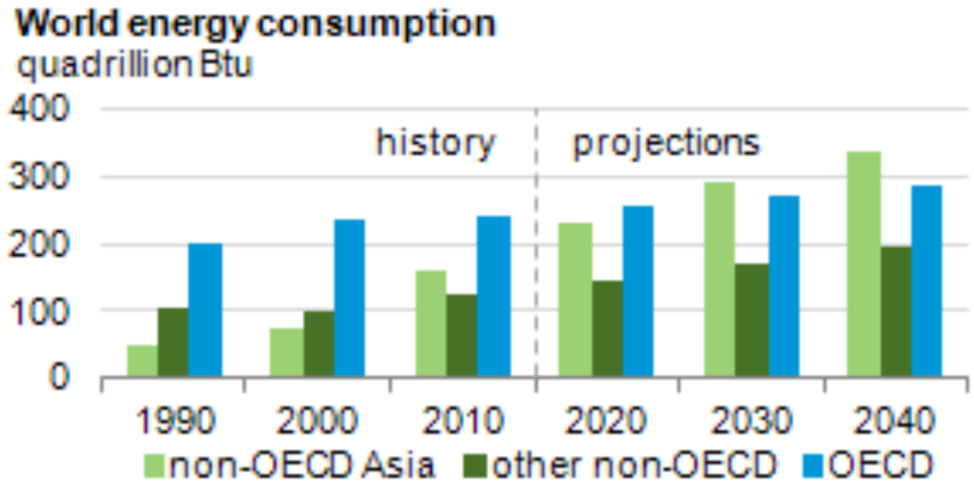
Energy consumption

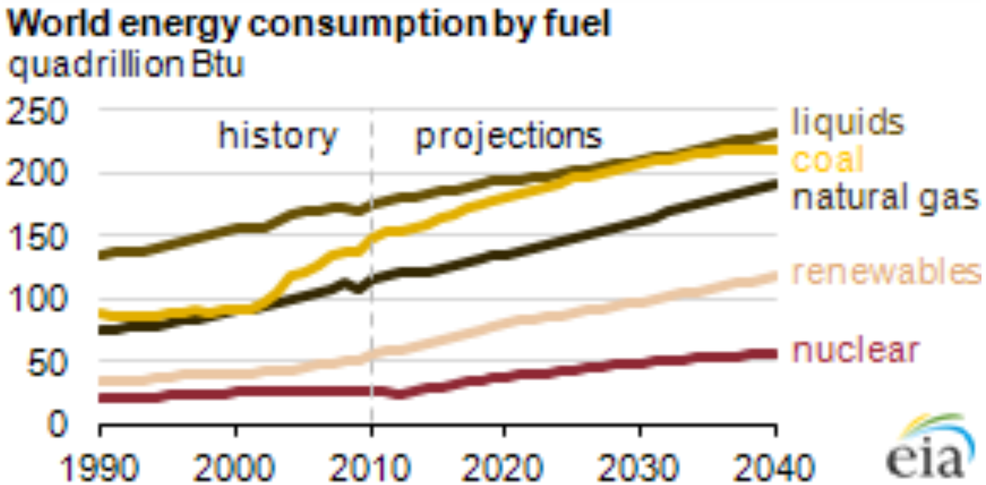


World energy consumption







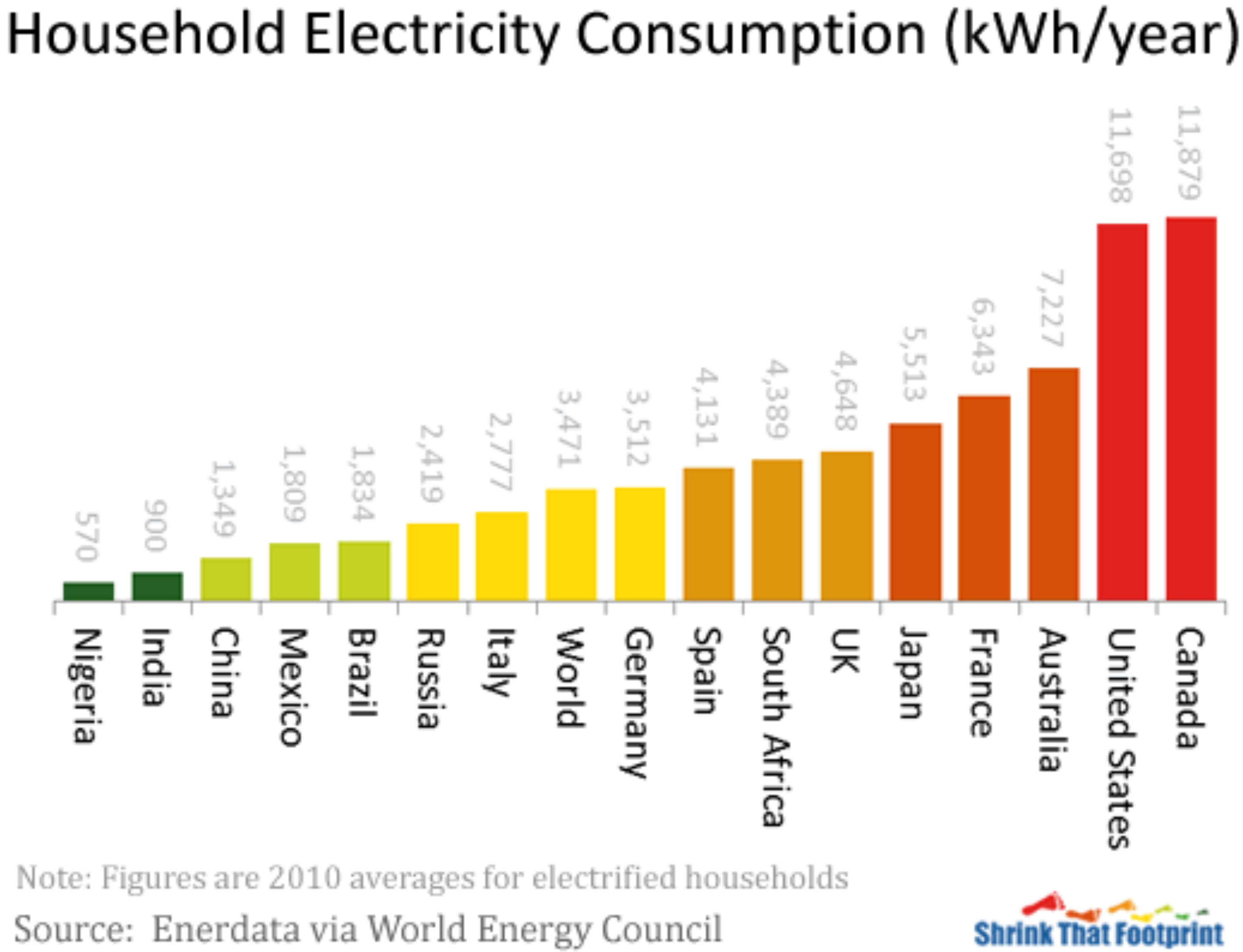


Source: www.eia.gov

World nightlight

Source: Max Dannenbaum/Stone/Getty Images

Electricity consumption



Word Cloud

- Which variables are likely to influence electricity demand?

Slido.com #57478

Overview

- Electricity demand
- Seasonality
- Weather dependence
- Long-term multivariate forecasts
- Density forecasts
- Univariate short-term forecasts

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

Quiz

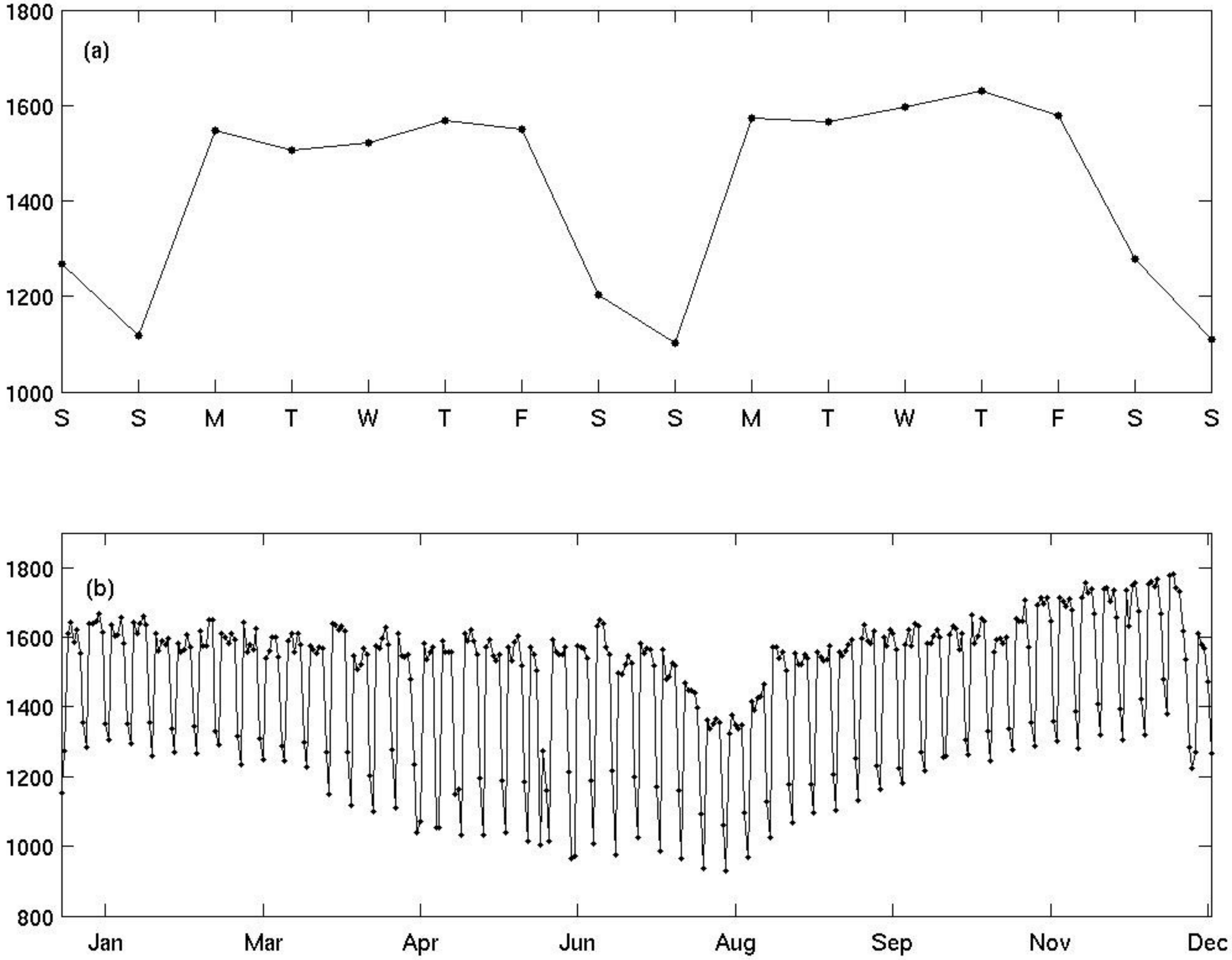
- When forecasting electricity demand, weather data is most important for which forecast horizon?
  a) 1 hour
  b) 6 hours
  c) 1 day
  d) 1 week

- **Slido.com #57478**

Forecasting models

- Require different models for different forecast lead times
- Are weather forecasts available and reliable?
- Univariate for short-term (< 1 day)
- Multivariate with weather variables included for medium-term
- Multivariate with simulated weather for long-term

Demand seasonality



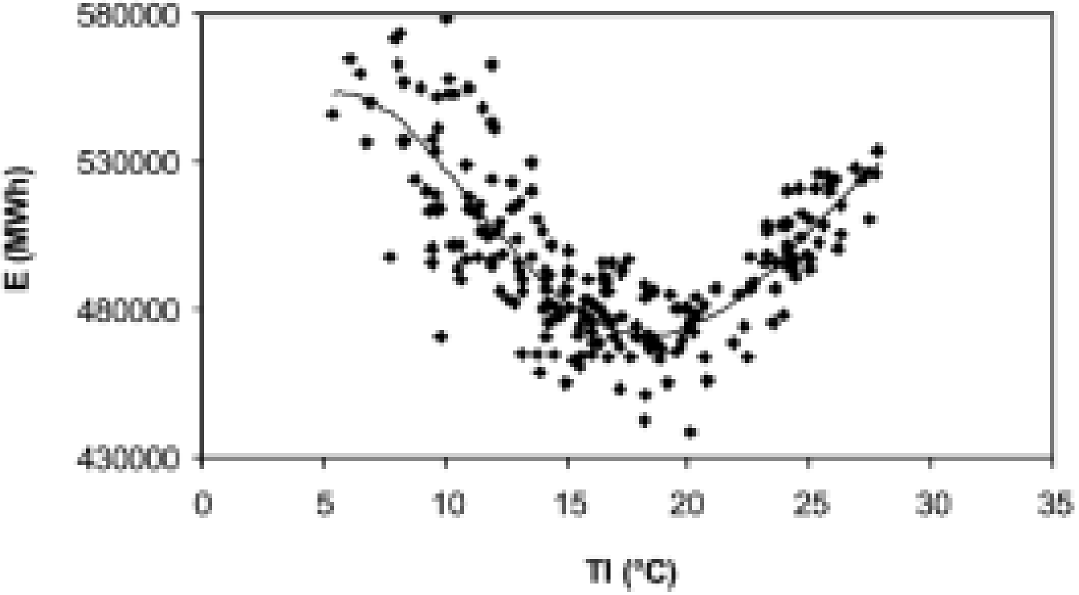
Influences on demand

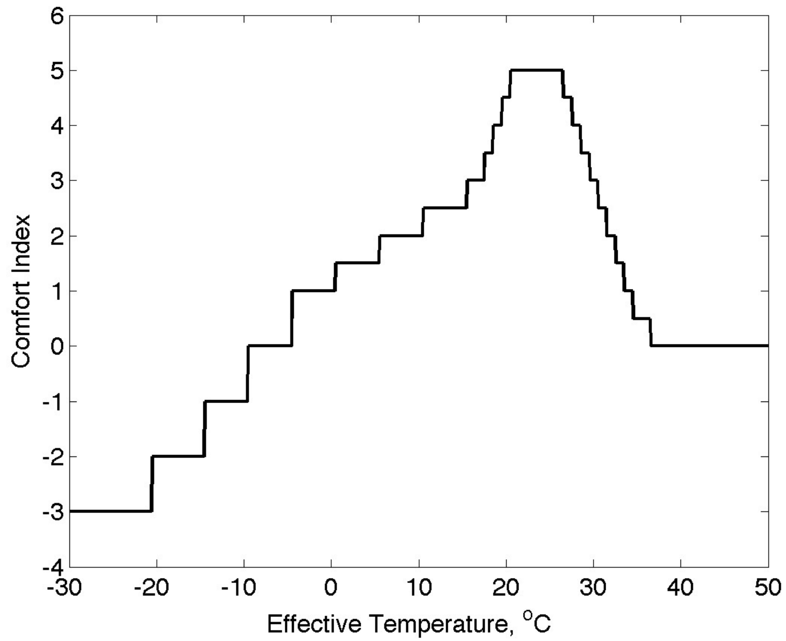
- Deterministic seasonality:
  - daily and weekly seasonality (low demand at weekends)
  - Annual seasonality (low demand during the summer)
- Calendar effects:
  - Summer holidays, Christmas holidays
  - Bank holidays
  - Bridge days, elections, strikes, …
- Weather effects:
  - Temperature
  - Wind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

Nonlinear dependence on temperature

- Demand tends to increase with decreasing temperature due to the use of heating equipment
- Demand also tends to increase with increasing temperature due to the use of cooling equipment such as air conditioners
- For the Netherlands the optimal temperature giving rise to the lowest demand was 15^o C

Temperature indices





- Heating and cooling degree days:
- $HDD_t = \max(18 - T_t, 0)$ $CDD_t = \max(T_t - 18, 0)$
- Effective temperature combines temperature and humidity

Simpson & McSharry (2011). An economic assessment of the impacts of climate change on tourism in Barbados. UN-ECLAC

Cooling power

- Cooling power is defined to describe the influence of draught on demand using a nonlinear function of temperature and wind speed
- Cooling power, C_t, on day t:

$$C_t = \begin{cases} W_t^{1/2}(18.3 - T_t) & \text{if } T_t < 18.3^o C \\ 0 & \text{if } T_t \geq 18.3^o C \end{cases}$$

where W_t is the wind speed and T_t is the temperature on day t.

The demand model

- Demand, D_t , on day t:

$$
\begin{aligned}
D_t \quad = \quad & a_0 + a_1 t + a_2 t^2 + \sum_{i=1}^{3} b_i \delta_{t,i} \\
& + \sum_{i=1}^{4} \alpha_i \tau^i + \sum_{i=1}^{3} \delta_{t,i} \sum_{j=1}^{4} \beta_{ij} \tau^j \\
& + \sum_{k=1}^{11} \gamma_k \delta_{t,\mathrm{feast_k}} + \gamma_{12} \delta_{t,\mathrm{summer}} + \gamma_{13} \delta_{t,\mathrm{christmas}} \\
& + c_1 T_t + c_2 {T_t}^2 + c_3 C_t + c_4 I_t + \varepsilon_t. \qquad (1)
\end{aligned}
$$

where τ is a time of year variable giving the deterministic seasonality, I_t is the luminosity and the δ are the dummy variables.

Model details

- Demand growth: quadratic dependence on time
- Fourth order polynomial used to describe the deterministic annual seasonality
- Separate descriptions (levels and seasonality) for Friday, Saturday and Sunday
- Nonlinear weather dependence on temperature, cooling power and luminosity

Structural uncertainty

- Model inadequacy
- Model mis-specification
- A priori structural bias of the user:
  - "I like NNs"
  - "I already have code for ____"
  - "____ worked last time"
  - "____ are universal approximators"

Forecasts

- For each year between 1995 and 2000
- The four previous years were used to fit the model parameters
- Out-of-sample predictions were generated for that year
- These year ahead forecasts start from the 31st of December of the previous year

Evaluation statistics

- Normalised Root Mean Square Error (NRMSE)

$$NRMSE = \frac{\sqrt{\left\langle (\hat{D}_t - D_t)^2 \right\rangle}}{\sigma}$$

  where σ is the standard deviation of the demand

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \left\langle 100 \left| \frac{\hat{D}_t - D_t}{D_t} \right| \right\rangle$$

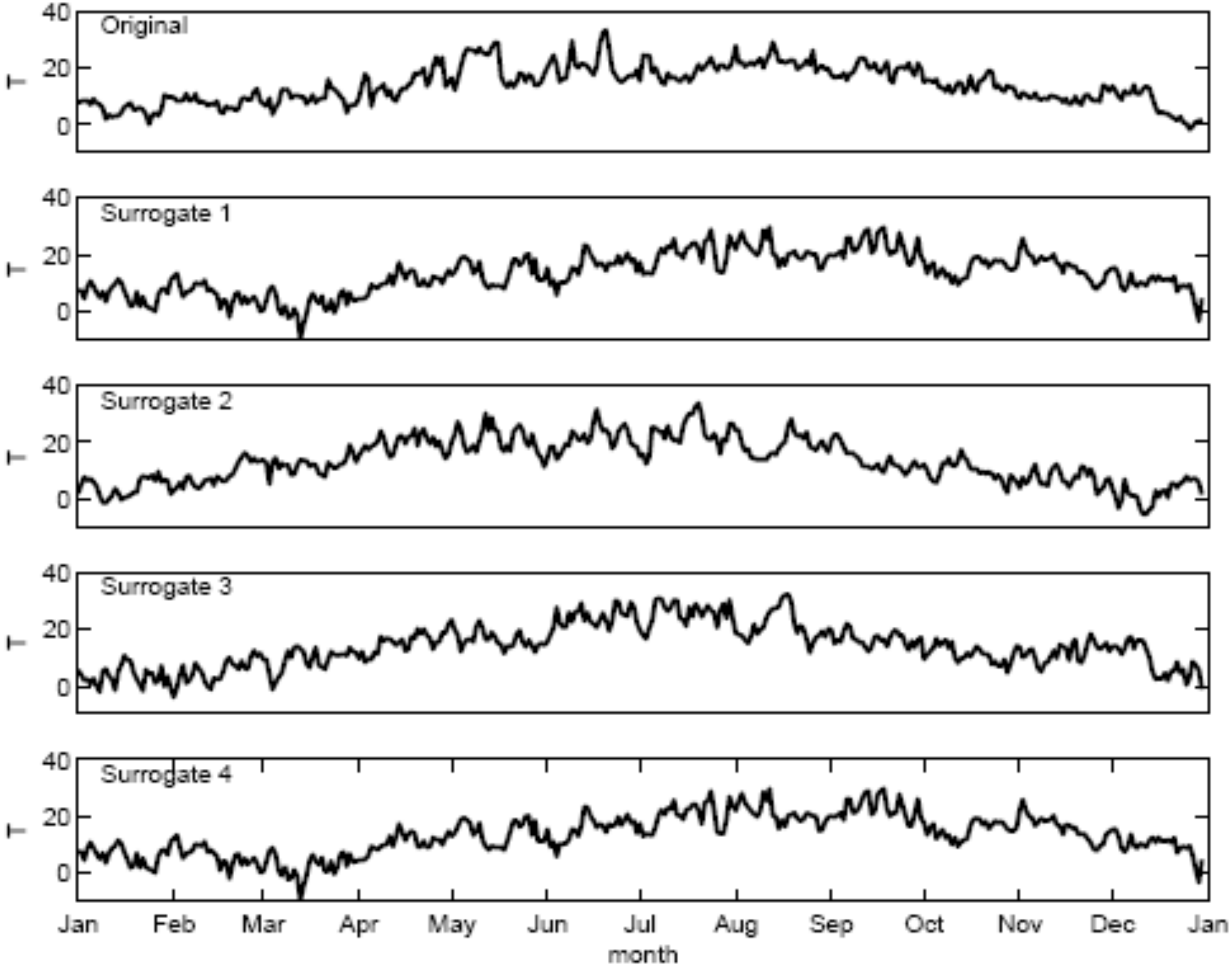
Future weather

- Weather forecasts are only of value out to two weeks ahead
- Require weather values for one year ahead
- Need to preserve linear autocorrelations and cross-correlations between the weather variables
- Also require similar histogram of values

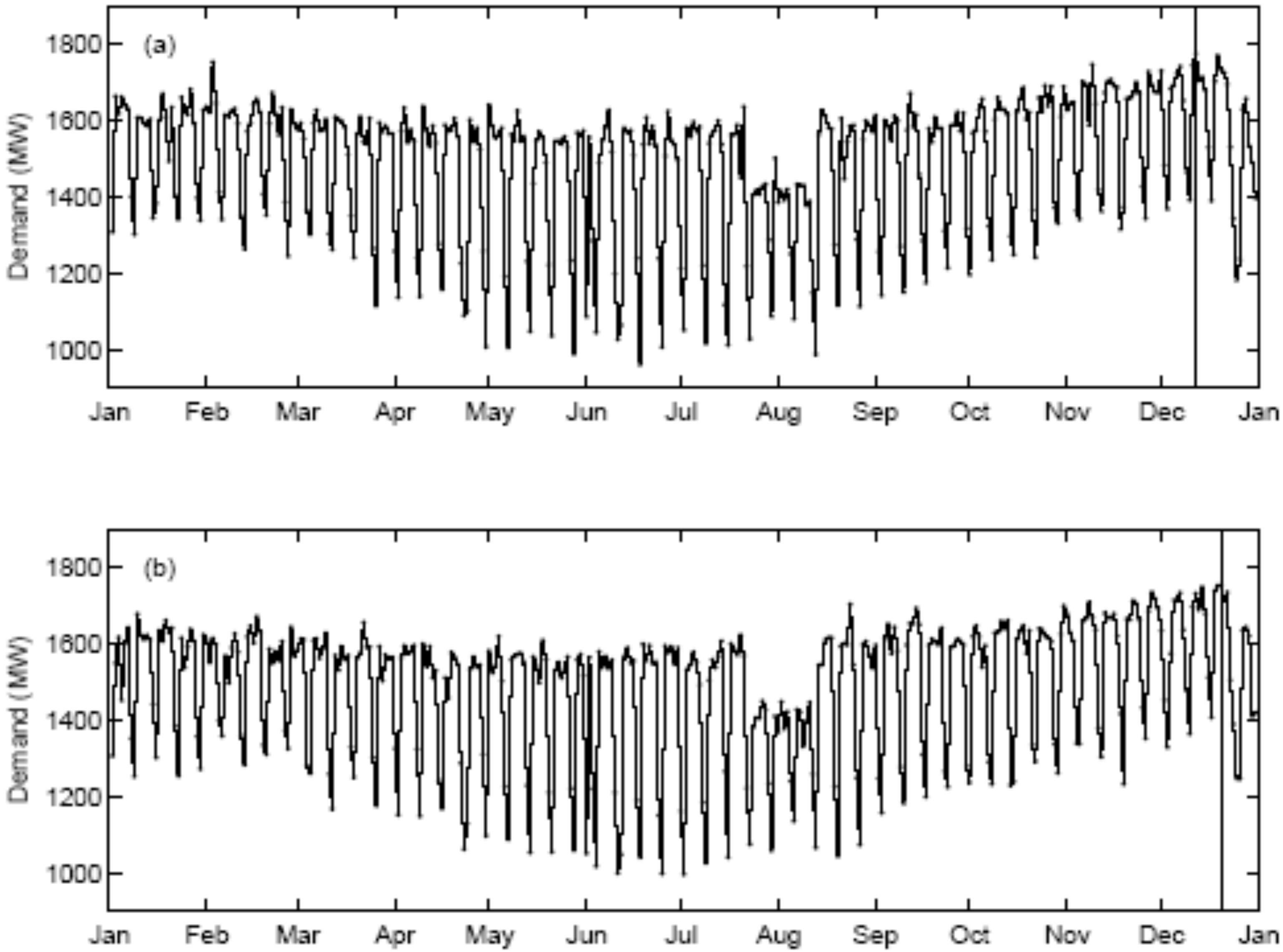
Weather simulation

- Distribution is maintained by shuffling the original time series
- Autocorrelation is preserved by fixing the amplitudes of the power spectrum
- Calculate the Fast Fourier Transform (FFT)
- Fix the amplitudes and scramble the phases (containing the nonlinear information)
- Transform back to the time domain
- Each replicate is known as a surrogate

Temperature surrogates



Year ahead demand forecasts



Probabilistic forecasting

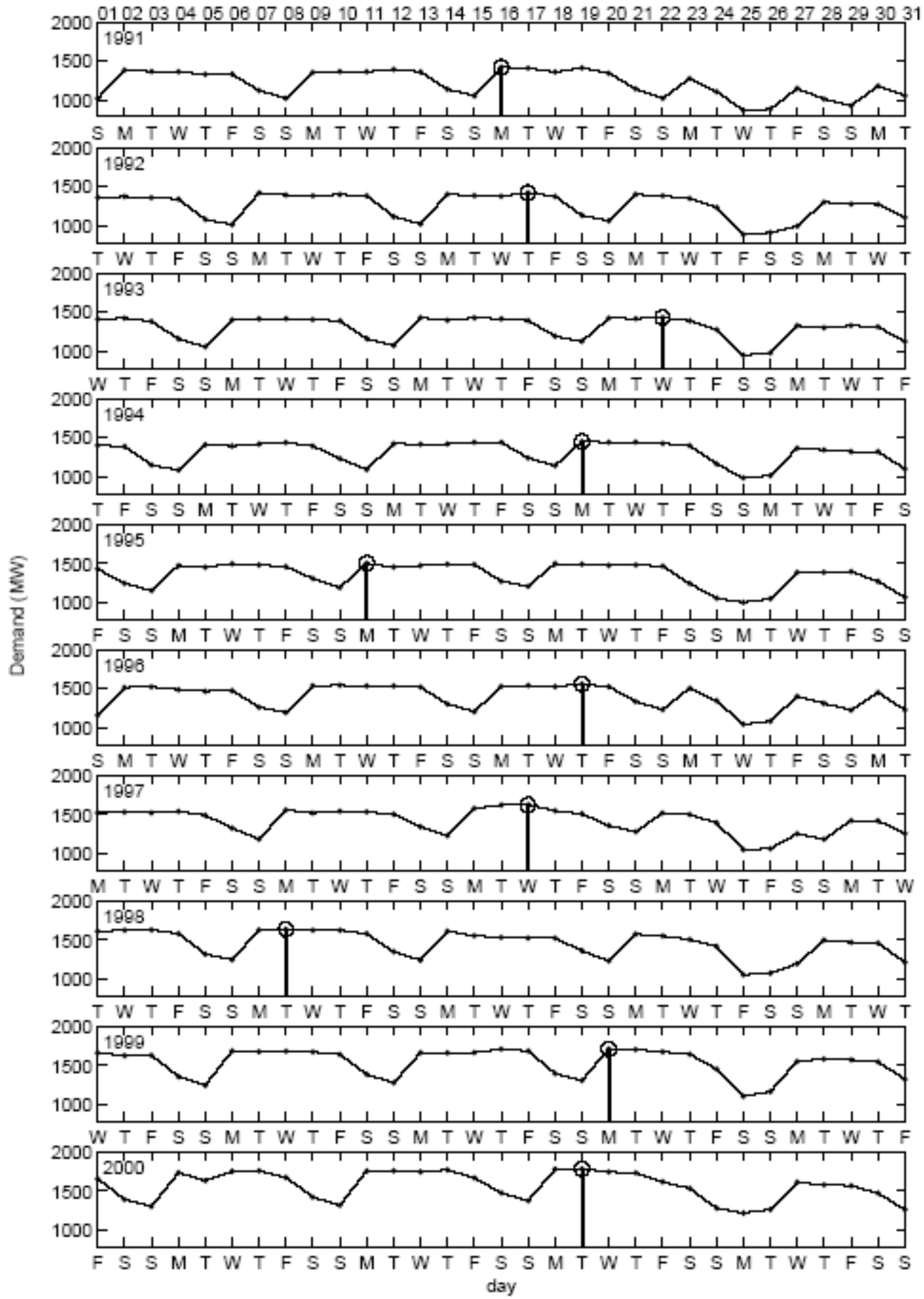
- Account for uncertainty:
  - Model uncertainty
  - Weather uncertainty
- Run 10000 Monte Carlo simulations
- Different weather scenarios
- Different error realisations
- Calculate the forecast PDF

Forecast results

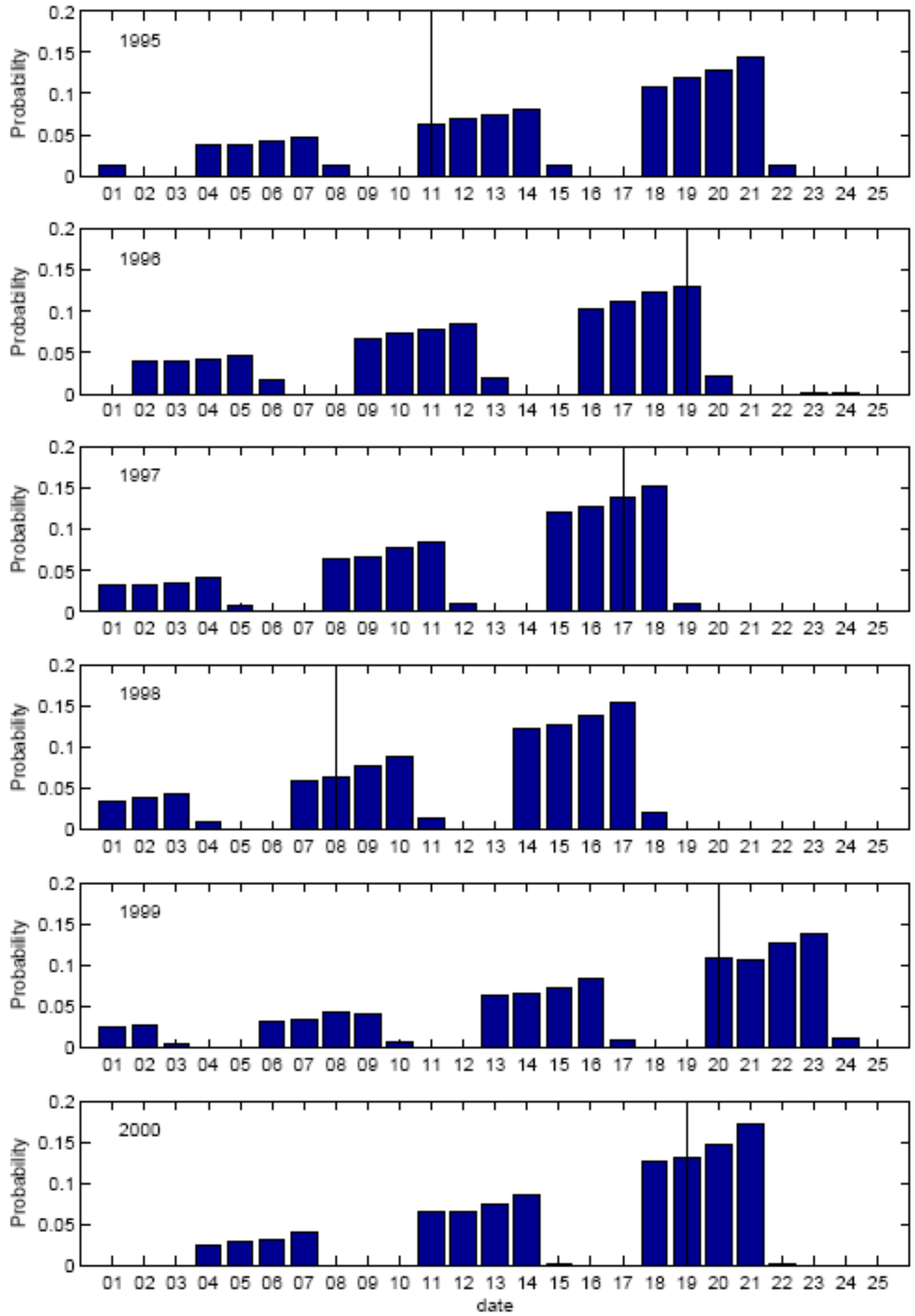
Forecast	Training		Testing		Testing	
Weather	Actual		Actual		Simulated	
Year	NRMSE	MAPE	NRMSE	MAPE	NRMSE	MAPE
1995	0.17	1.87	0.25	2.25	0.27	2.39
1996	0.20	1.87	0.23	2.56	0.22	2.39
1997	0.22	1.87	0.22	2.26	0.24	2.55
1998	0.22	1.93	0.23	2.48	0.25	2.75
1999	0.20	1.89	0.22	2.49	0.23	2.52
2000	0.19	1.90	0.30	3.11	0.27	2.81
Average	0.20	1.89	0.24	2.52	0.25	2.57

Peak demand

- Generally occurs on a weekday: Monday through Thursday
- Usually around the middle of December during the pre-Christmas rush

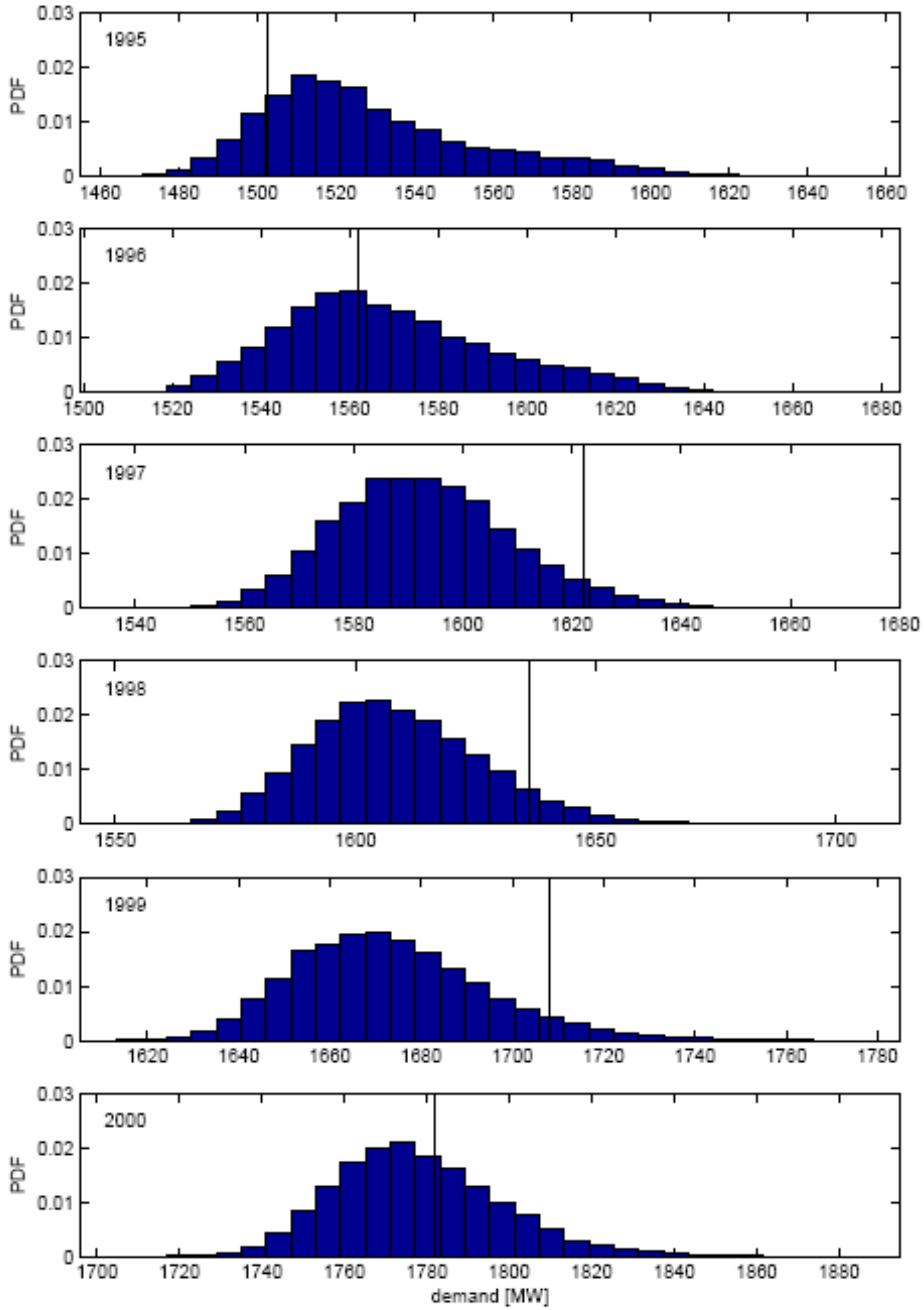


PDF Forecasts of peak timing



PDF forecasts of peak magnitude

Non-normal asymmetric forecast PDFs



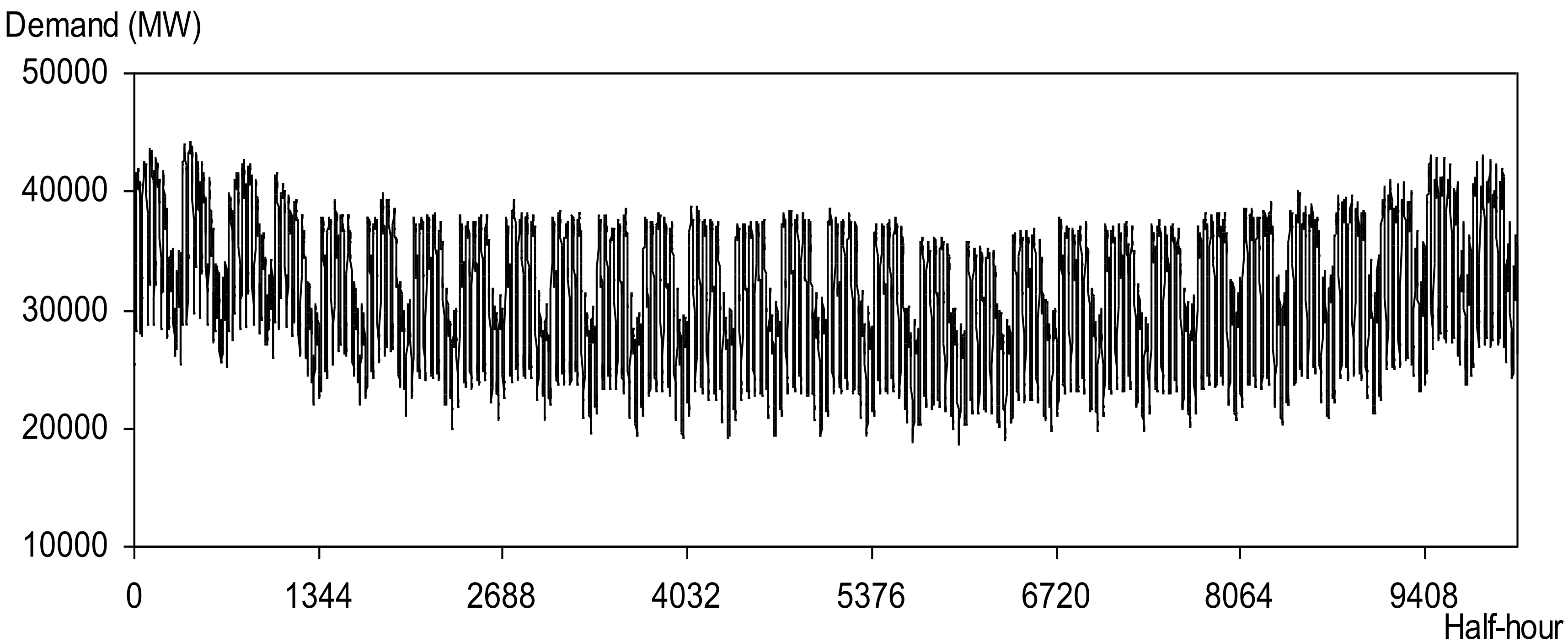
Poll

- For hourly demand data from customers, we generally expect to detect two cycles with periods (in hours) of:
  a) 1 and 7
  b) 7 and 24
  c) 24 and 168
  d) 24 and 365
- **Slido.com #57478**

Short-term (< 1 day) forecasts

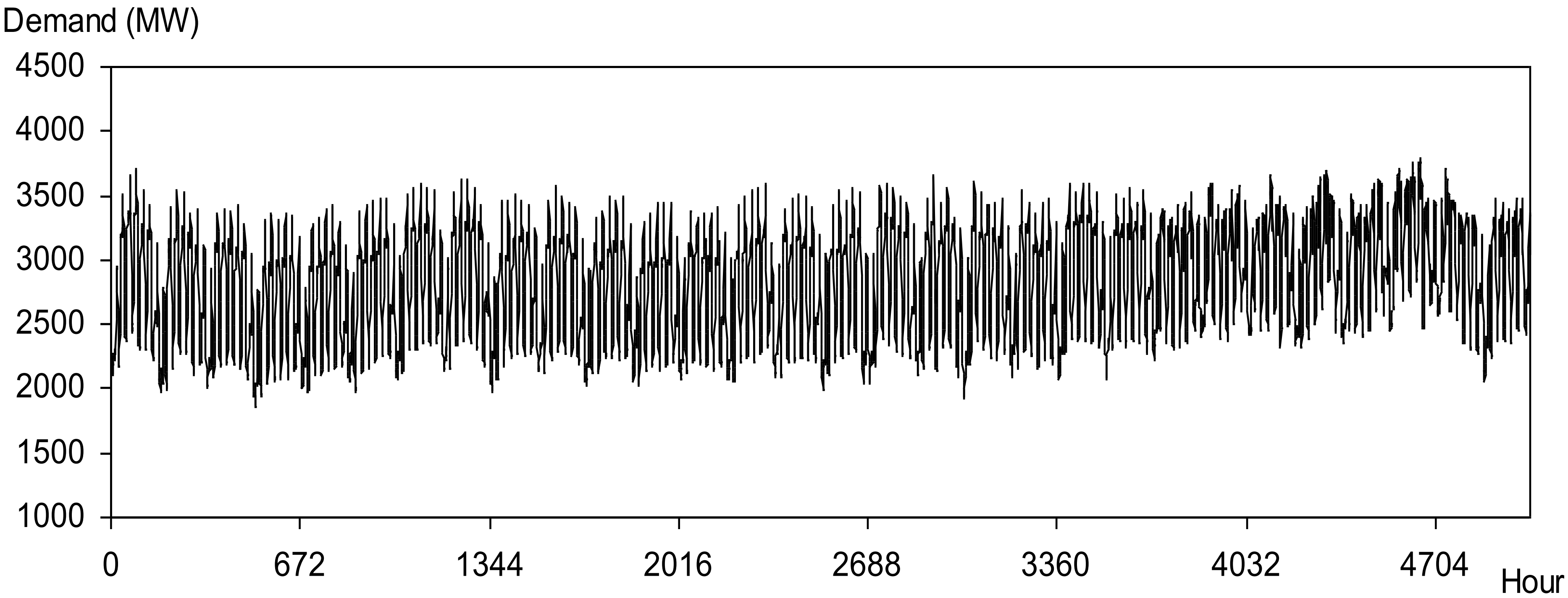
- Compare and evaluate four univariate methods – multiple hypotheses approach
- Use intra-day data (30 weeks) for the state of Rio in Brazil (hourly) and for England and Wales (half-hourly).
    - Rio: 3,360 observations for estimation and 1,680 for evaluation.
    - England and Wales: 6,720 observations for estimation and 3,360 for evaluation.

England and Wales series: within-day cycle (48 periods) and a within-week cycle (336 periods).



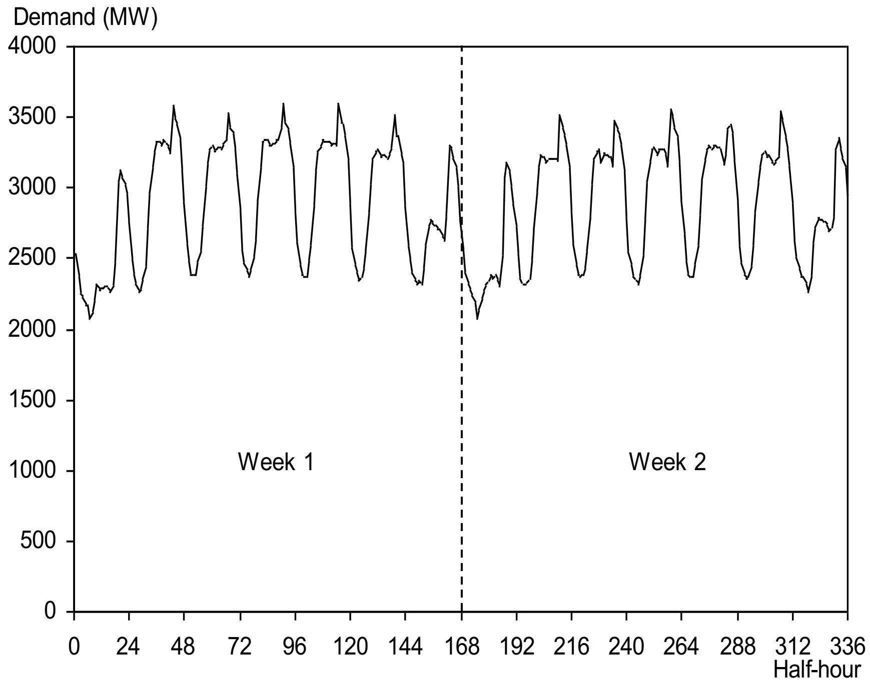
Monday 27 March 2000 to Sunday 22 October 2000

Rio series: within-day cycle (24 periods), within-week cycle (168).

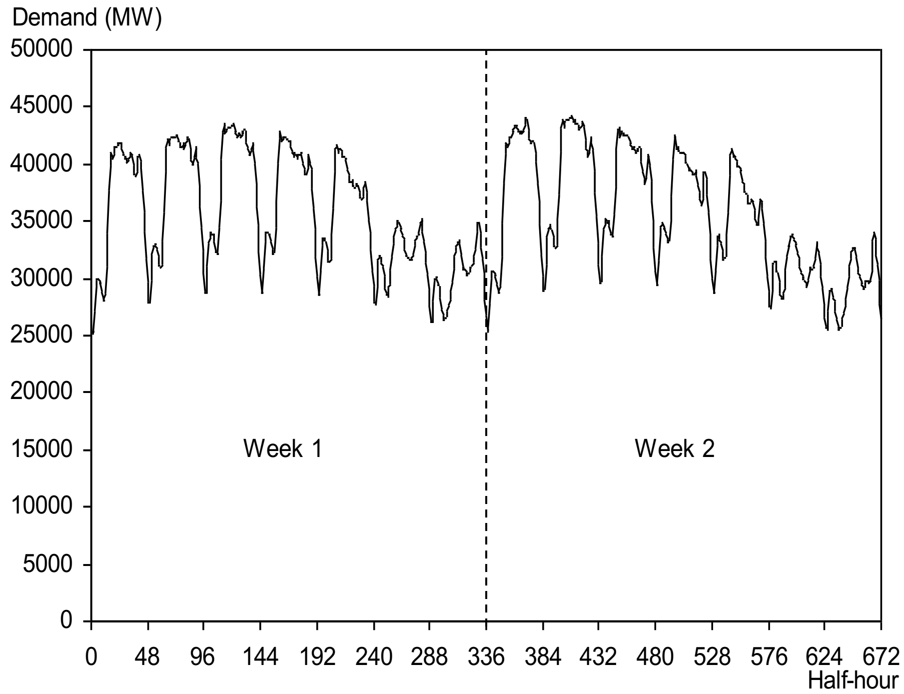


Sunday 5 May 1996 to Saturday 30 November 1996

Two weeks of data...



Rio



England & Wales

Methods

(1) Box and Jenkins SARIMA-good track record;

(2) an exponential smoothing alternative - robust, attractive for online demand forecasting (Taylor, 2003);

(3) an artificial neural network implementation that performed well on similar data (Darbellay and Slama, 2000);

(4) a principal component analysis approach - combines decomposition and regression, but reduces # of possible regressions: focus on PCs.

Double Seasonal ARIMA

$$(p,d,q)\times(P_1,D_1,Q_1)_{s_1}\times(P_2,D_2,Q_2)_{s_2}$$

- Standard multiplicative ARIMA:

$$\phi_p(L)\ \Phi_P\left(L^s\right)\nabla^d\ \nabla_s^D\ X_t = \theta_q(L)\ \Theta_Q\left(L^s\right)\varepsilon_t$$

- extended to model two seasonalities in the

demand (s_1= 24, 48 and s_2= 168, 336) :

$$\phi_p(L)\,\Phi_{P_1}\left(L^{s_1}\right)\Omega_{P_2}\left(L^{s_2}\right)\nabla^d\ \nabla_{s_1}^{D_1}\ \nabla_{s_2}^{D_2}\ X_t$$

$$= \theta_q(L)\,\Theta_{Q_1}\left(L^{s_1}\right)\Psi_{Q_2}\left(L^{s_2}\right)\varepsilon_t$$

The Double SARIMA models

- Selection Criteria: lowest SBC, residuals
  - Rio:
    - ARIMA$(3,0,3)\times(3,0,3)_{24}\times(3,0,3)_{168}$.
  - England and Wales:
    - ARIMA$(2,0,1)\times(2,0,1)_{48}\times(1,0,2)_{336}$

Double Seasonal Exponential Smoothing

- Similarly, extend Holt-Winters:

$$\hat{\boldsymbol{y}}_{\boldsymbol{t}}(\boldsymbol{k}) = (\boldsymbol{S}_{\boldsymbol{t}} + \boldsymbol{k}\,\boldsymbol{T}_{\boldsymbol{t}})\,\boldsymbol{D}_{\boldsymbol{t}-\boldsymbol{s}_{\boldsymbol{1}}+\boldsymbol{k}}\,\boldsymbol{W}_{\boldsymbol{t}-\boldsymbol{s}_{\boldsymbol{2}}+\boldsymbol{k}}$$

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

Double Seasonal Exponential Smoothing

- Accuracy improved by a adjusting 1st-order autocorrelation:
  - AR(1) model, $e_t = \lambda e_{t\text{-}1} + \xi_t$, 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 $\lambda^k e_\tau$.

Double Seasonal: 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

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

Artificial Neural Network

- Single hidden layer feed-forward network: output is demand and inputs (x_{it}) are lag demand:

$$f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w}) = g_2\left(\sum_{j=0}^{m} v_j \, g_1\left(\sum_{i=0}^{k} w_{ji} \, x_{it}\right)\right)$$

$g_1(\cdot)$ and $g_2(\cdot)$ are activation functions (sigmoidal and linear)
w_{ji} and v_j are the weights (parameters).

- Weights are estimated by:

$$\min_{\boldsymbol{v},\boldsymbol{w}}\left(\frac{1}{n}\sum_{t=1}^{n}\left(y_t - f(\boldsymbol{x}_t,\boldsymbol{v},\boldsymbol{w})\right)^2 \;+\; \lambda_1\sum_{j=0}^{m}\sum_{i=0}^{k} w_{ji}^2 \;+\; \lambda_2\sum_{j=0}^{m} v_i^2\right)$$

  – where n is the number of in-sample observations, and λ_1 and λ_2 are regularisation parameters that penalise complexity

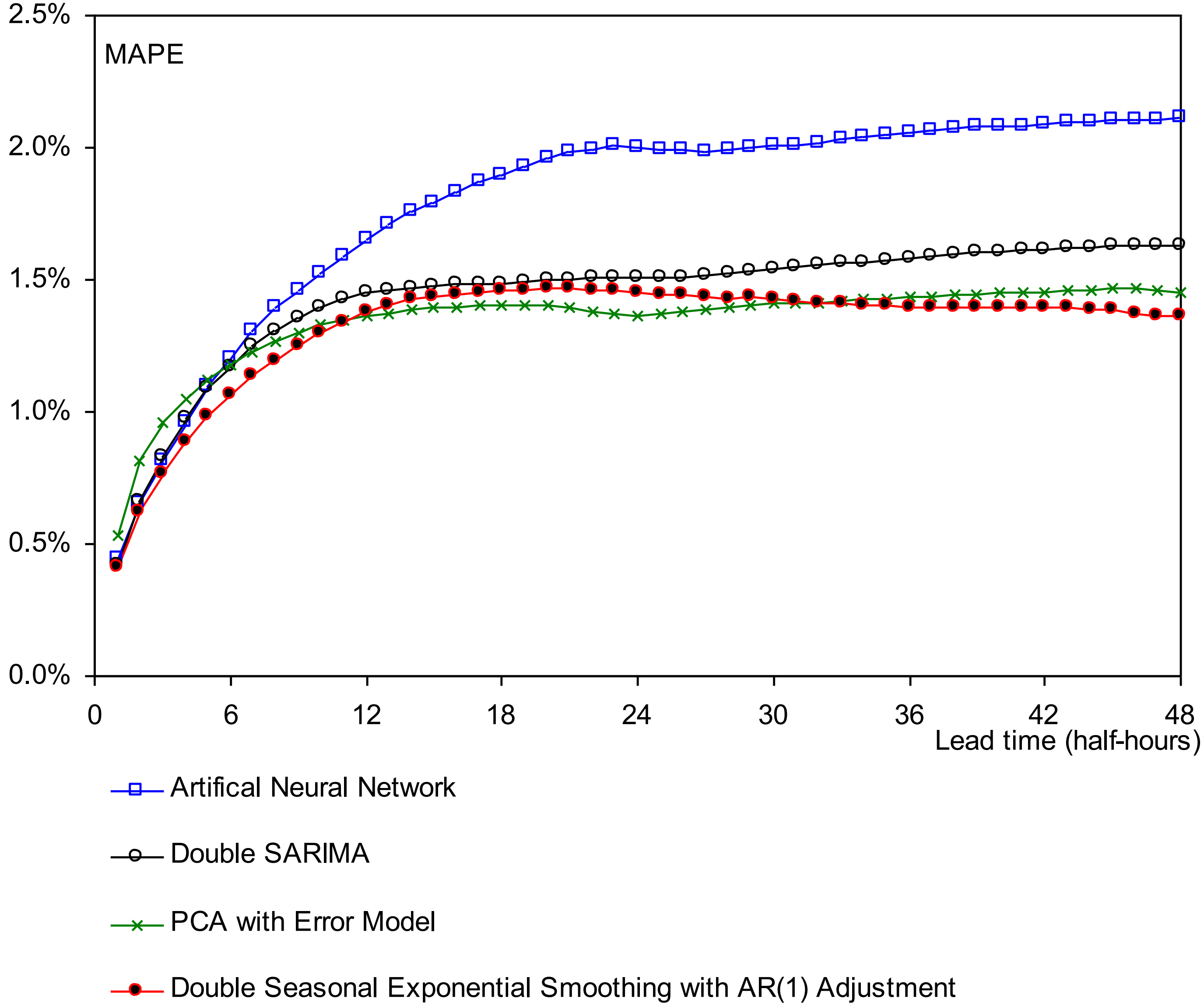
- Stationarity => seasonal and first differences

- Following the hold-out method:
  - *England &Wales:* $\lambda_1=\lambda_2=0.005$ and $m=14$, and the following lags: 1, 48, 49, 96, 144, 192, 240, 288, 336, 384, 432, 480, 624 and 672.
  - *Rio:* $\lambda_1=\lambda_2=0.005$ and $m=12$, and the following lags in our model: 1, 2, 24, 25, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264, 312 and 336.

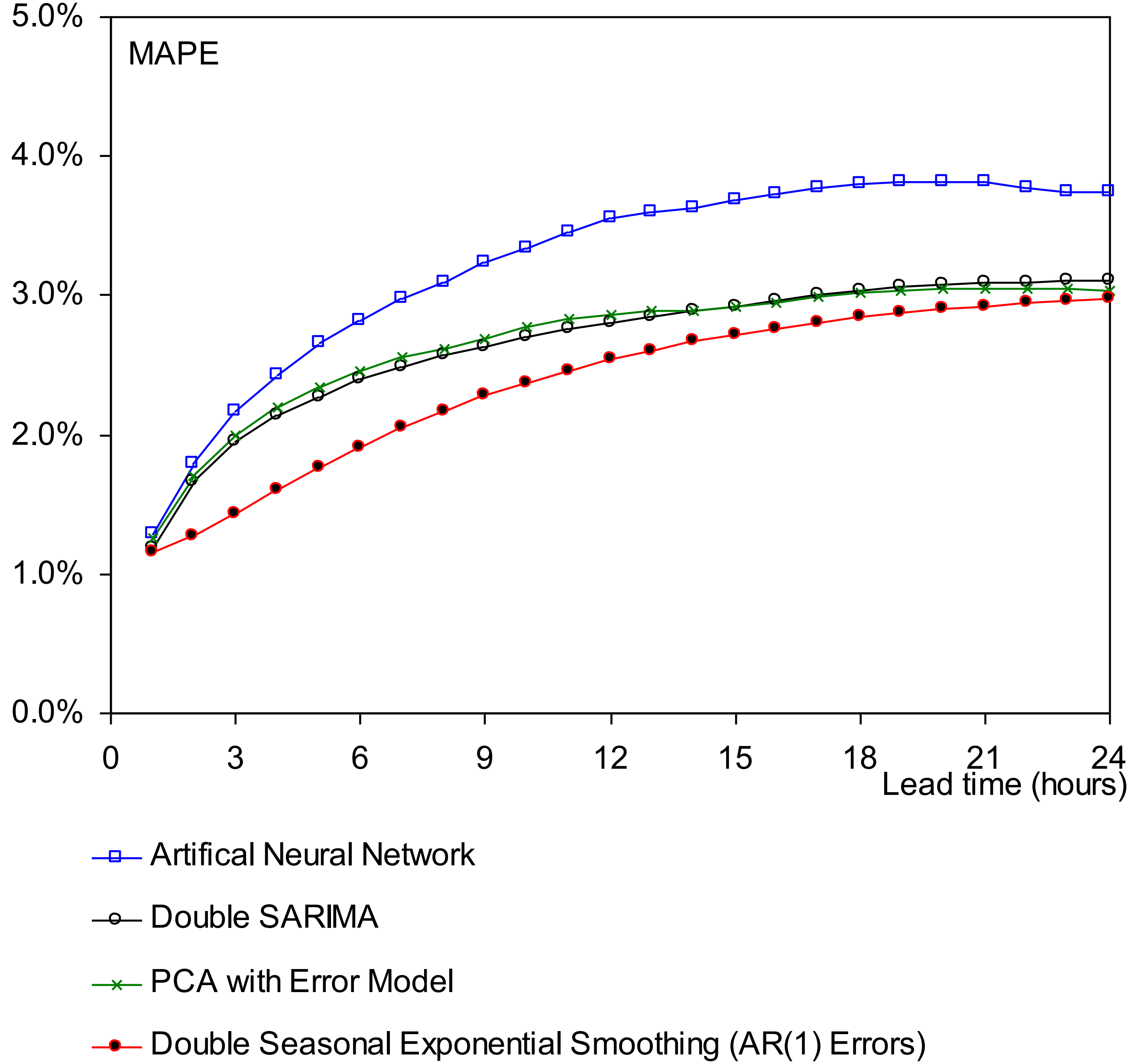
PCA Based Method

- PCA reduces the data and thus captures the essence of the day-to-day similarity
- Each PC is modelled for the weekly effect:
  - regressions with dummy variables for the weekday effect, linear and quadratic terms on t
- # of weeks to estimate the components and the # of components need to be decided
- serially correlated errors => model e(t)

MAPE: England & Wales 10-week post-sample



MAPE: Rio, 10-week post-sample



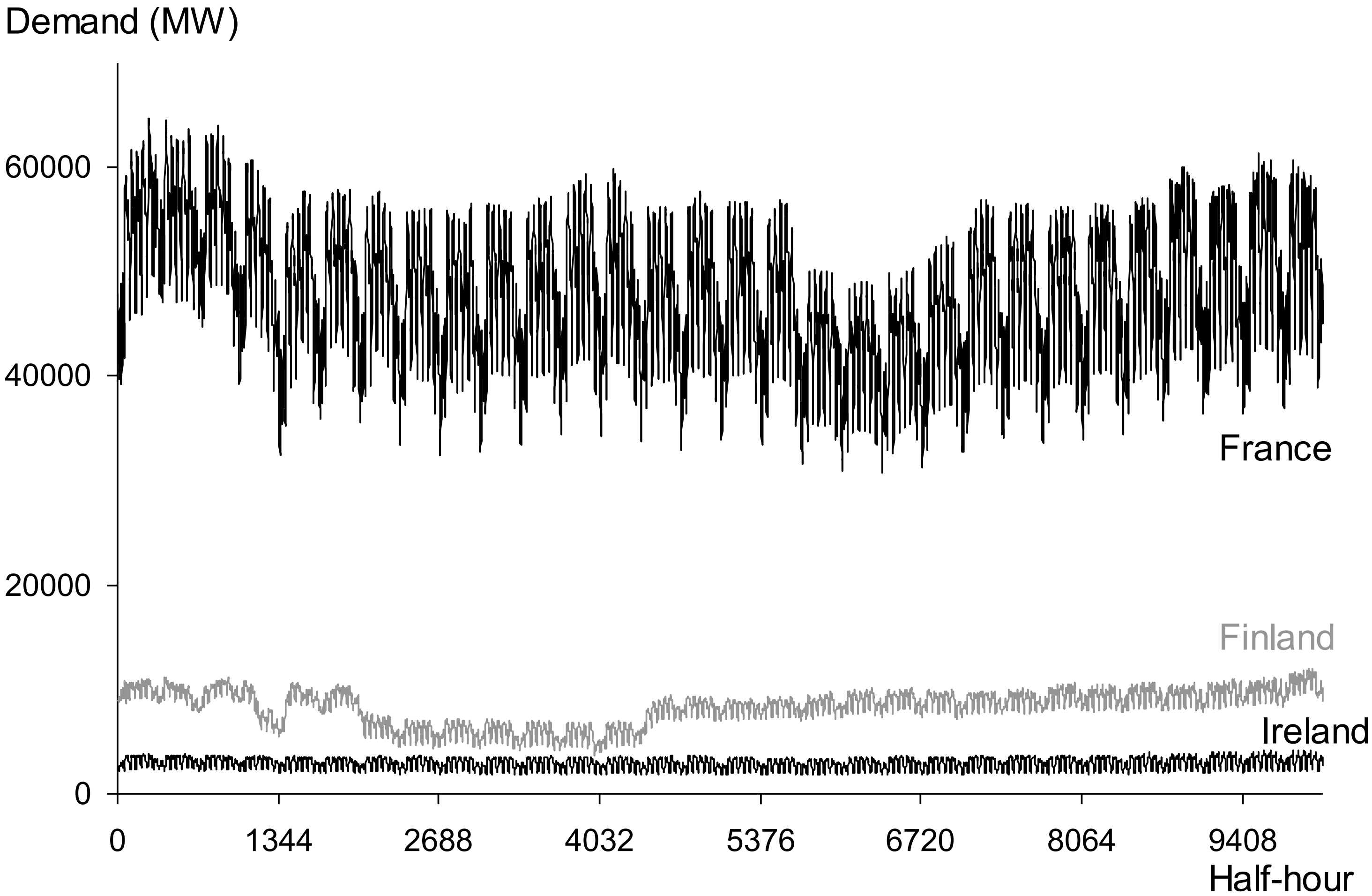
EU comparison

- Intraday electricity demand data from ten European countries
- Found presence of both an intraweek and an intraday seasonal cycle
- Similar ranking of methods across the ten time series
- Best performance from the double seasonal Holt-Winters exponential smoothing method.

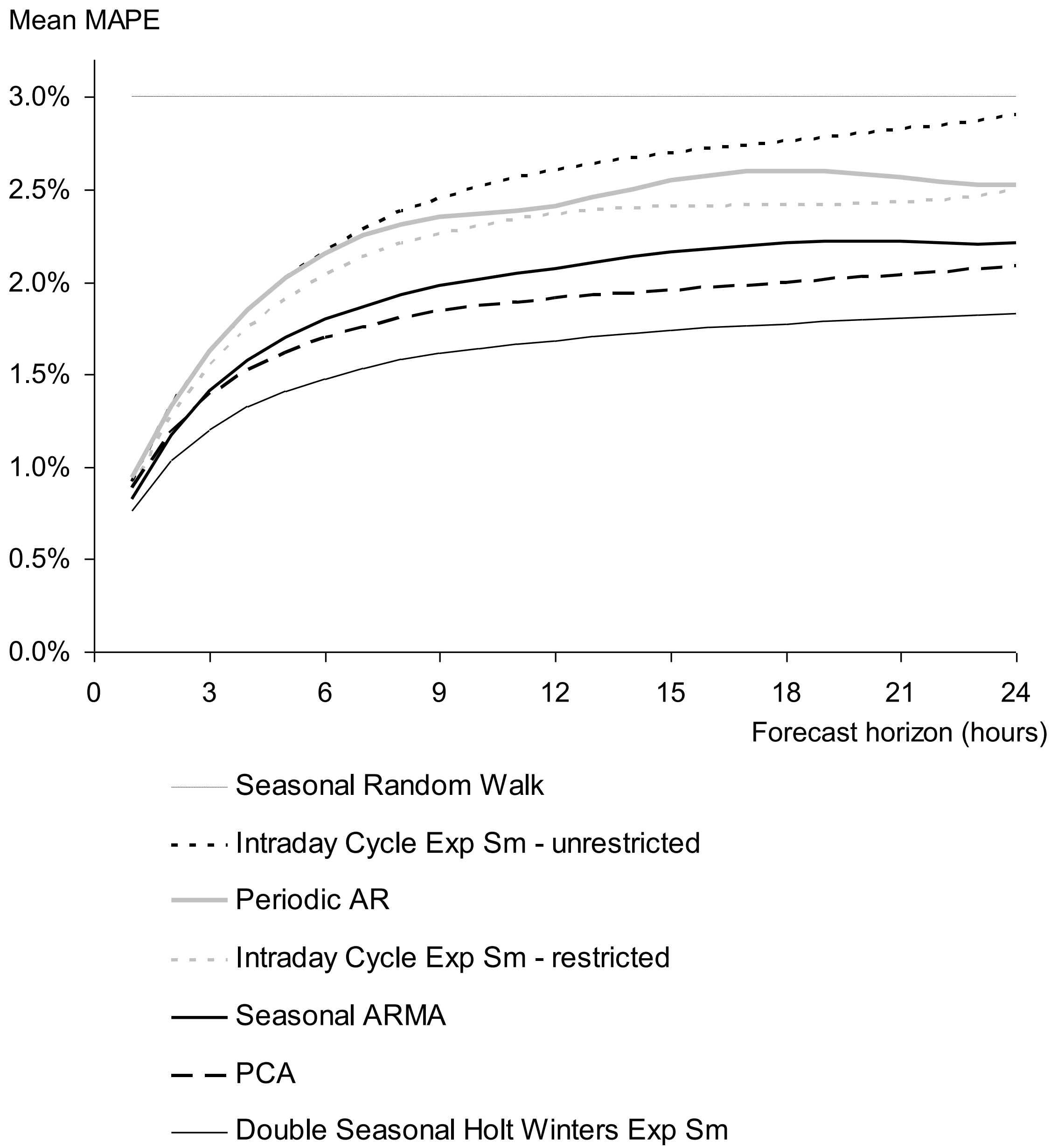
Mean Load and Population

	Mean Load (GW)	Population (Million)	Mean Load per Capita (W)
Half-hourly			
Belgium	9.4	10.4	906
Finland	8.3	5.2	1,588
France	47.8	60.9	786
Great Britain	36.0	58.9	611
Ireland	2.9	4.1	711
Portugal	5.2	10.6	492
Hourly			
Italy	32.7	58.1	563
Norway	12.0	4.6	2,599
Spain	25.5	40.4	630
Sweeden	14.5	9.0	1,608

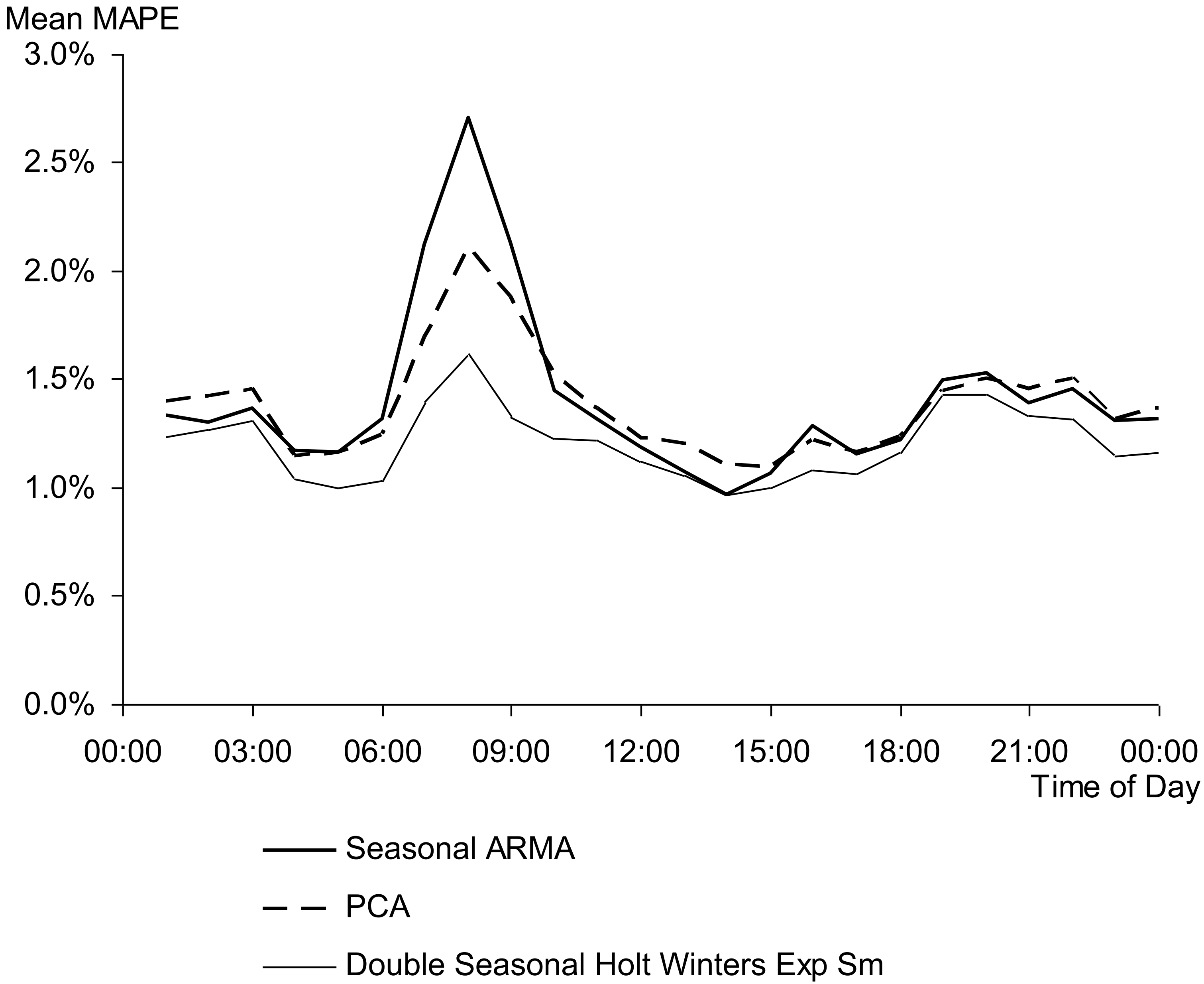
Electricity Demand



Forecast Performance



Intraday Performance



References

- McSharry, PE, Bouwman, S and Bloemhof, G. (2005) Probabilistic forecasts of the timing and magnitude of peak electricity demand. IEEE Transactions on Power Systems, 20(2): 1166-1172
- Taylor, JT, de Menezes, LM and McSharry, PE (2005) A comparison of methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting, 22(1): 1-16
- Taylor, JT and McSharry, PE (2007) Short-Term Load Forecasting Methods: An Evaluation Based on European Data. IEEE Transactions on Power Systems, 22(4): 2213 - 2219