THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

Phase identification and load forecasting with home energy data

Richard Bean

ITEE, University of Queensland

Centre for Energy Data Innovation

The **Centre for Energy Data Innovation** is a research enterprise created in partnership between Redback Technologies and The University of Queensland. The centre is partnering with Redback Technologies, Energy Queensland, the Springfield City Group, Microsoft and Advance Queensland to unlock new insights into electricity networks using IoT technology at a neighbourhood level.

The research centre is focused on data aggregation and analytics in the field of power and energy systems. The integrating of meters, digital controls and cloud computing will enable new data products for direct use by multiple disparate industries. Research, development, and utilization of these technologies will profoundly change the status quo, and see new business models, products and services in sectors such as insurance, finance, transport, infrastructure planning, home automation, and energy networks.

The centre focus will be the research and development of three principal technologies to transform energy grids and energy consumption. They are:

1. **Embedded Networks & Transactional Systems**
2. **Human-centred design of interactive Information Visualisations**
3. **Data Science, Big Data Analytics, and Machine Learning**

Some milestones of the Centre

Under the **Data Science** milestones, developed in conjunction with Redback and Energy Queensland

- Forecasting (aggregate) home energy data at feeder level and above
  - short and medium term
- Phase identification of houses on a feeder
- Non-intrusive load monitoring (NILM)
- Detecting broken neutral events; or, more generally, anomaly detection
- Understanding network state – graph visualization

Motivations for forecasting and phase ID

- AEMO/DNSP average and peak load forecasts at 10, 50, 90% probabilities of exceedence (PoE) in SoO/APR/APS docs; for simulations leading to business decisions; reserve level studies

Short term forecasting (e.g. hour to day ahead)

- Point forecasting versus probabilistic forecasting
- Virtual power plant, embedded networks
- Battery scheduling – minimizing costs by reducing grid imports at peak times (time of use tariffs)

Medium term forecasting (e.g. month ahead)

- Switching feeders/maintenance
- Based on ensemble weather forecasts (NWP)

Phase identification

- Poor record keeping – home medical device / life support devices
- Large fines - $20,000 – to be avoided
- Phase balancing – reducing losses

Redback Ouija Lite devices

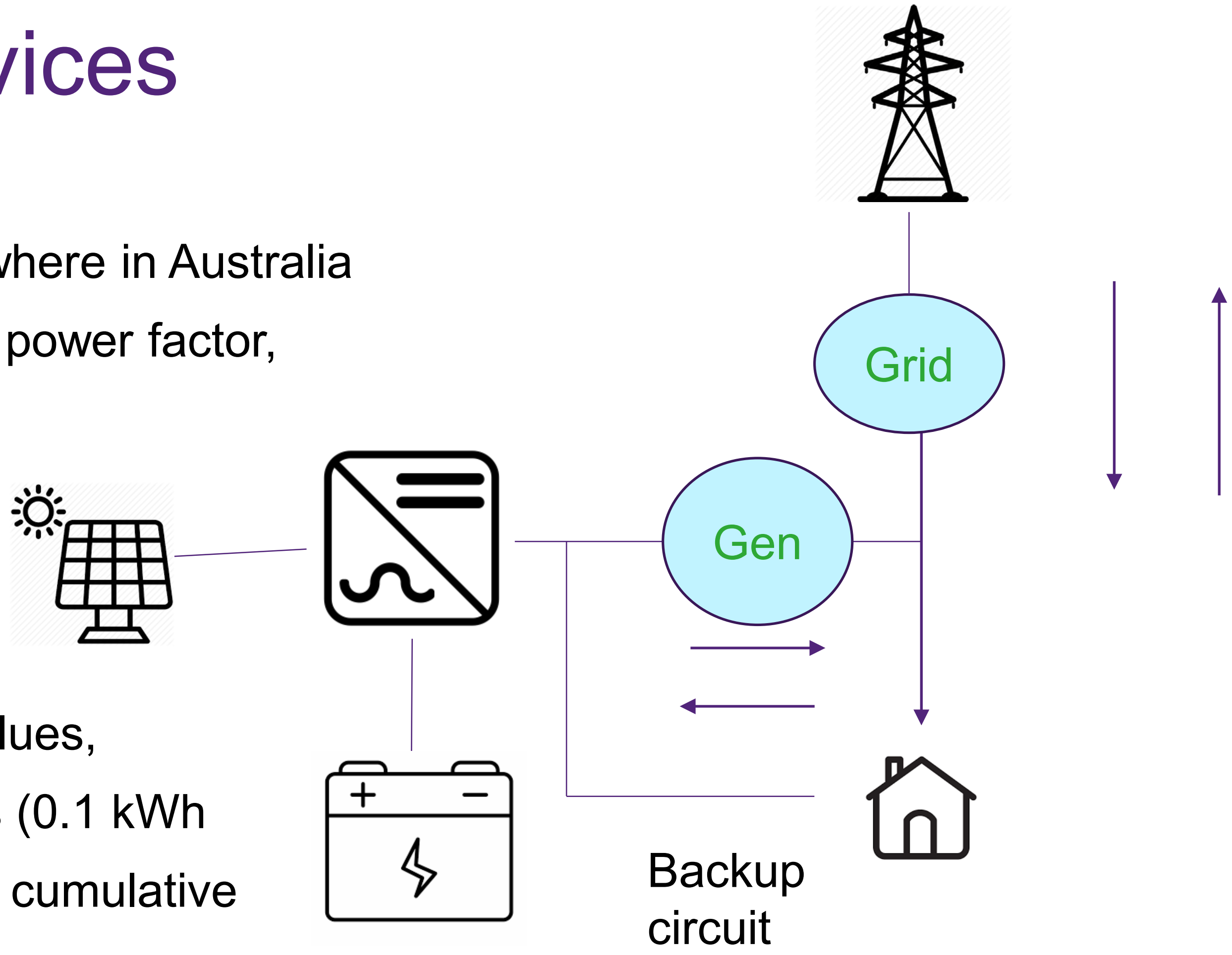
What devices record

How devices differ from smart meters elsewhere in Australia

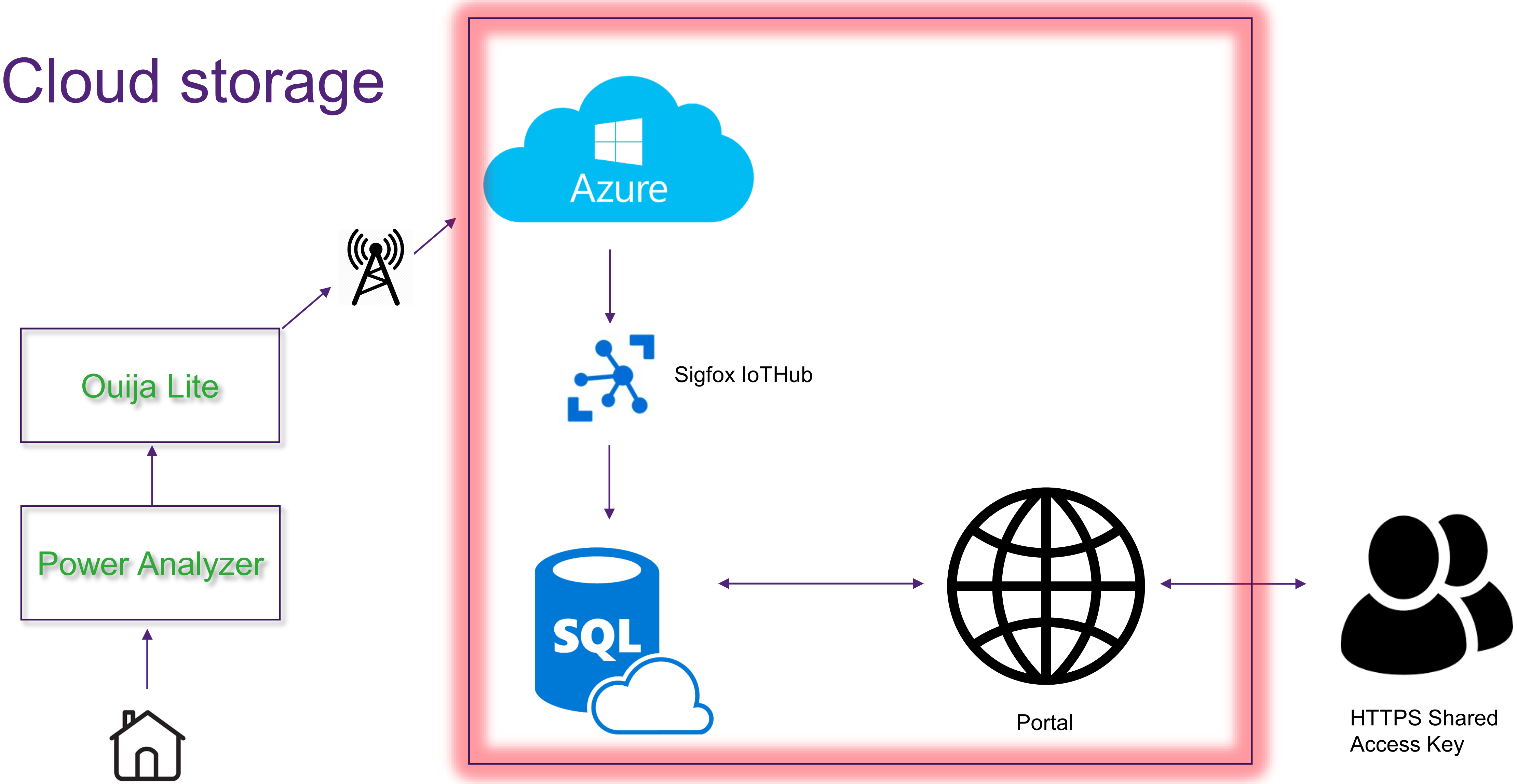
Recording grid voltage, generation current, power factor, impedance, grid import/export energy, generation import/export energy.

Instantaneous values, or cumulative values for energy.

One minute resolution for instantaneous values, cumulative values updated every two hours (0.1 kWh resolution). But diversity in summing house cumulative values.



Cloud storage



Phase identification

Prior research

Pezeshki and Wolf (2012) – correlation based method – “sag and swell”

Using voltage data from houses and transformer

Supervised clustering – three phases, plus transformer data to identify which phase each house was on

Further research – confidence levels around phase ID; costs of phase balancing

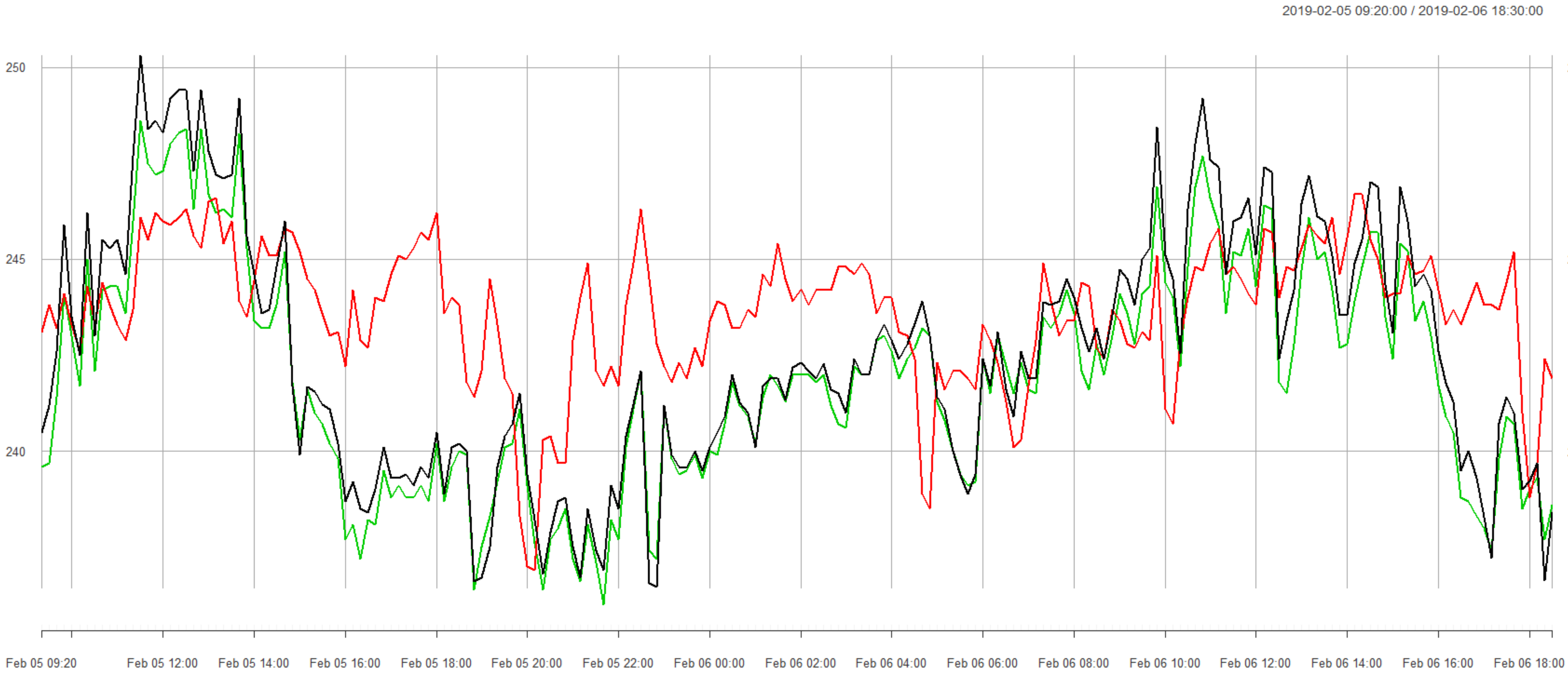
Limited data – good data hard to come by – P & W had 74 houses

Legal obstacles – privacy considerations – addresses/NMIs as identifiers

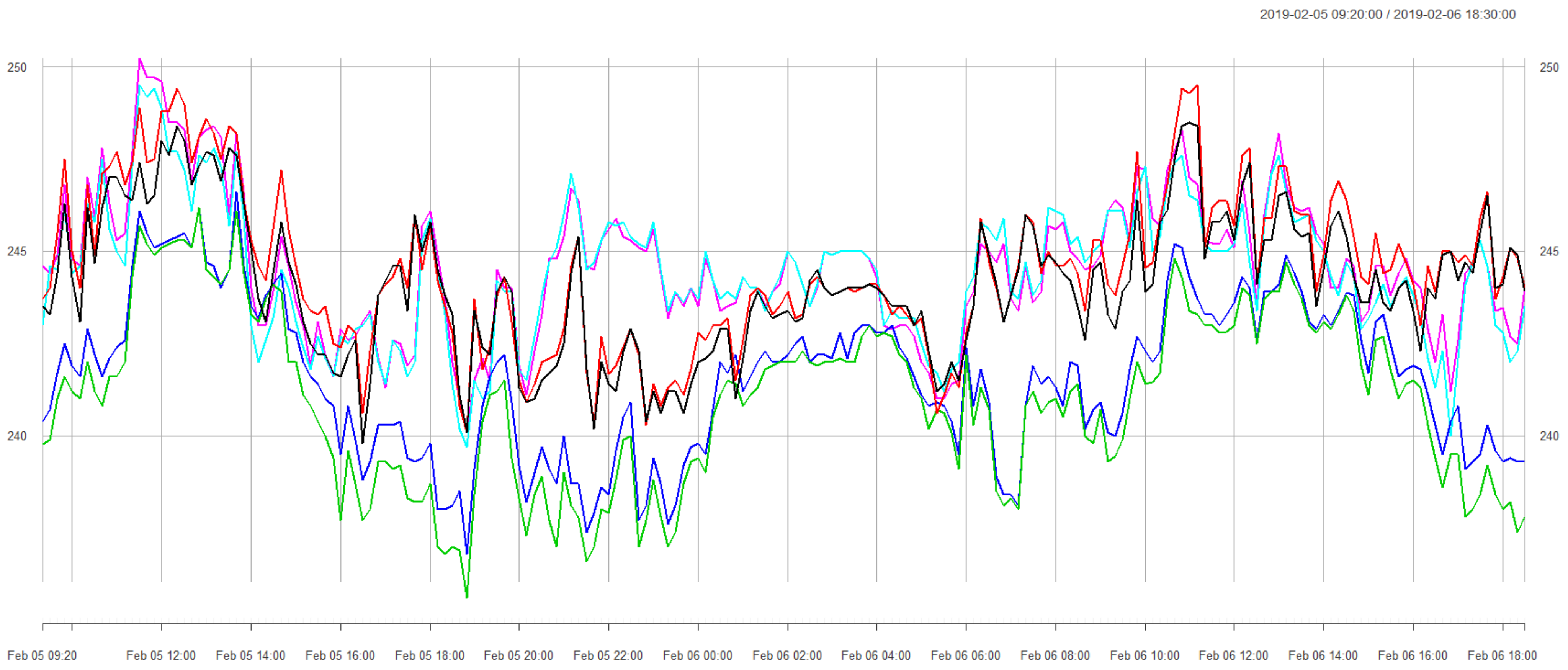
Unsupervised clustering – partitioning around medoids (PAM – Kaufman and Rousseuw 1987)

Can’t identify order of phases (road to house) – more research needed – but aids DNSP to correct existing records

Phase ID time series – 10-minute aggregation



Phase ID time series – 10-minute aggregation



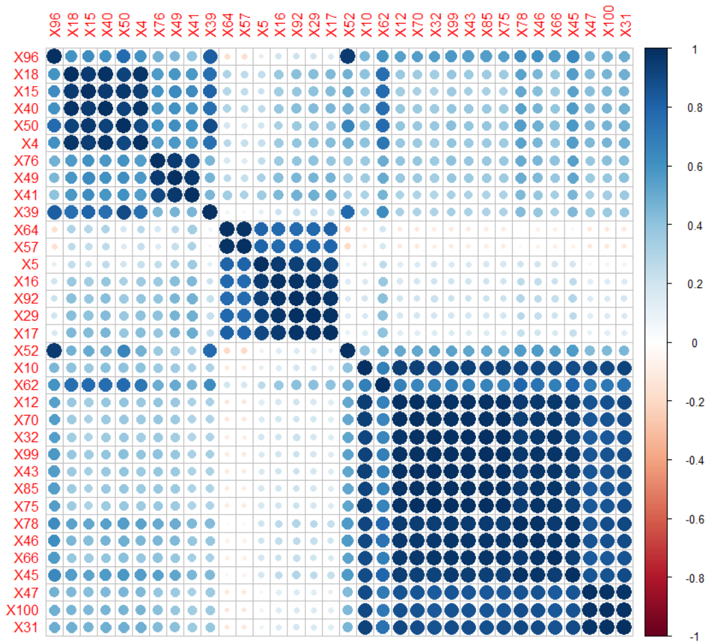
Phase identification

Aggregation choice – e.g. 10 minutes

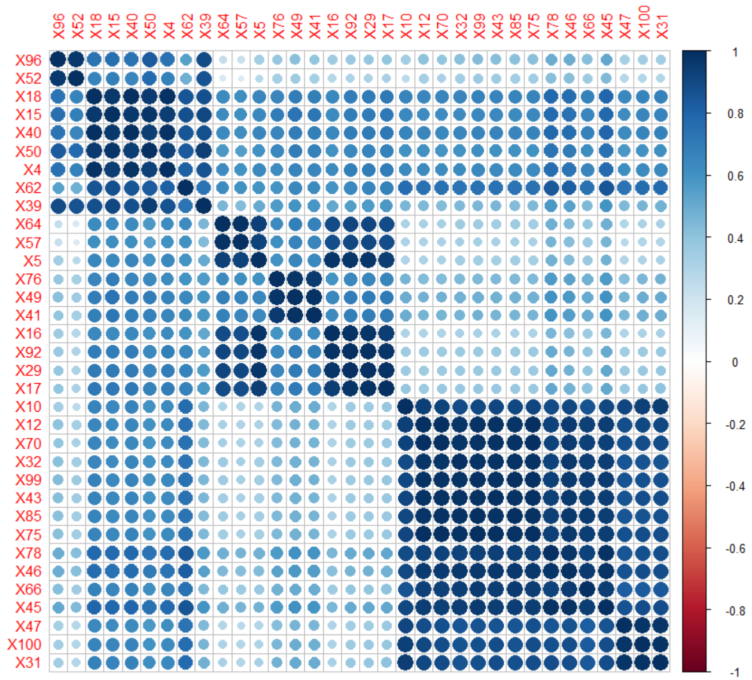
Subsetting time

Correlation plot

24 hour data – T1

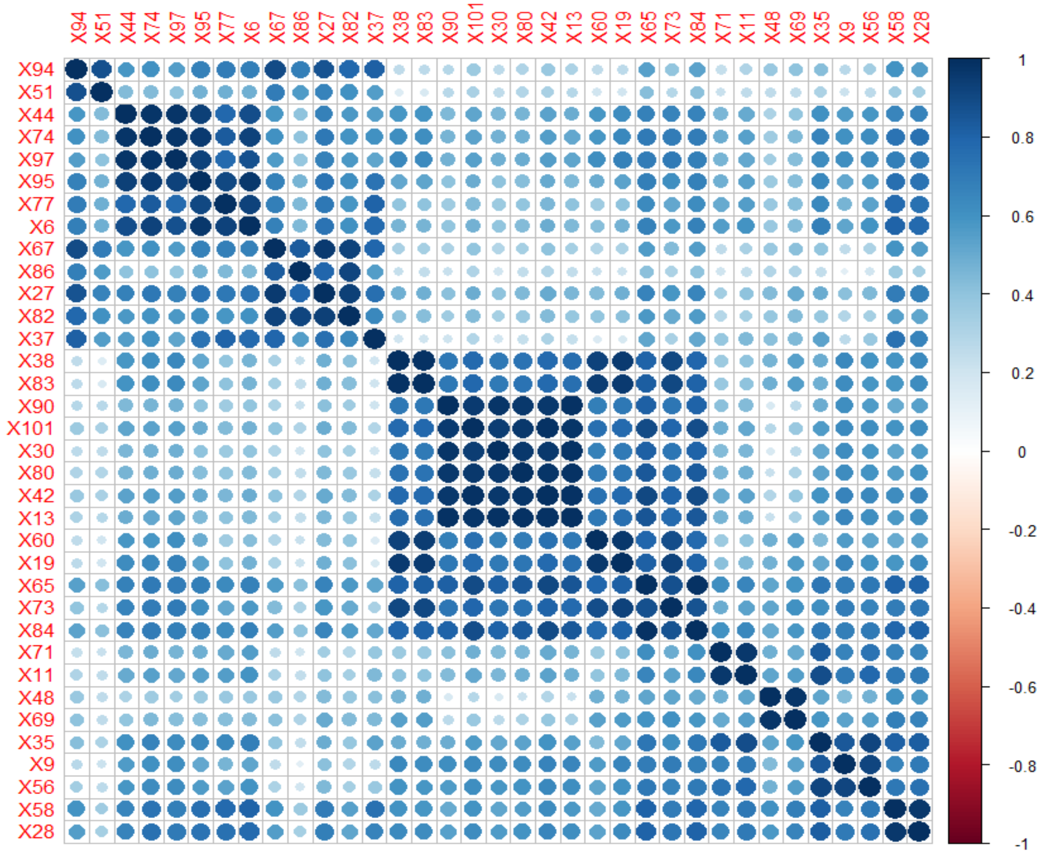


Time slice 3-6 am – T1

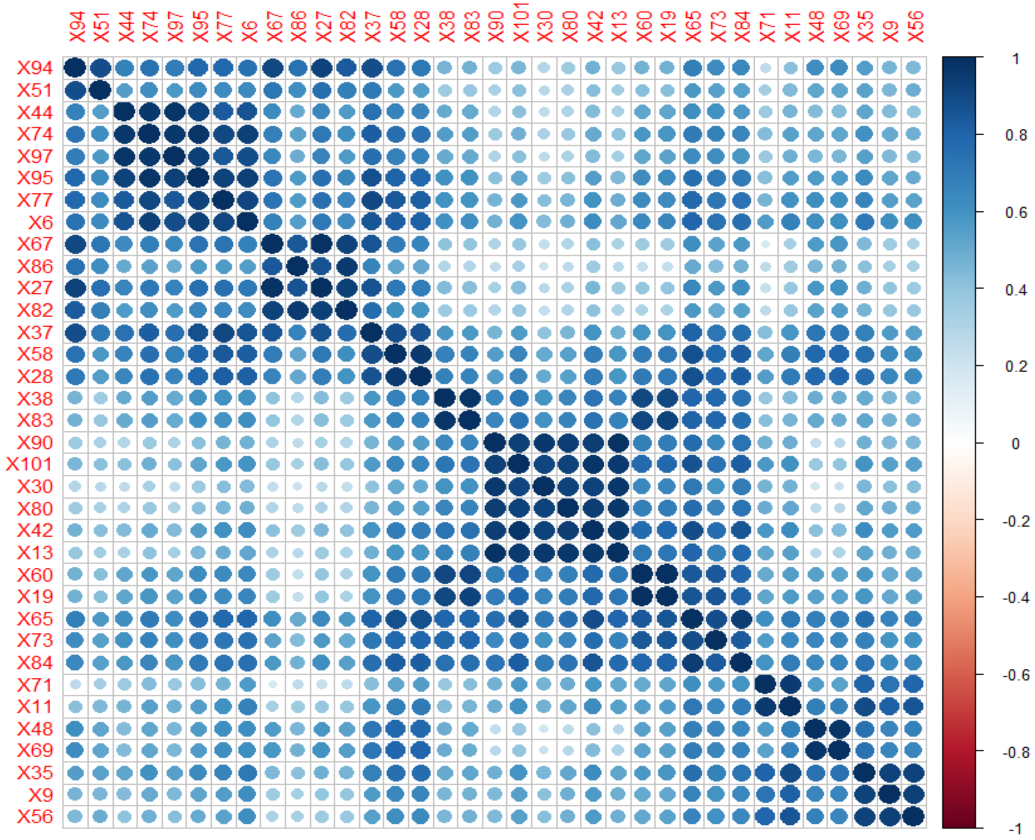


Phase identification

24 hour data – T2



Time slice 3-6 am – T2



Forecasting for short and medium term

Previous short term forecasting approach at Redback – for battery scheduling

Based on Global Energy Forecasting Competition (GEFcom) winning approaches – probabilistic forecasting of load, price, wind and solar (Hong et al 2016)

“Several teams in GEFCom2014 won multiple tracks using similar methodologies.” (Hong et al)

Quantile regression forest (Meinshausen 2006) using weather data from European Centre

At Redback - GFS data used – bounding box script; R, Azure, VMs, C# for GRIB parsing

Forecasting **house load** and **solar generation** (e.g. Solcast for short-term)

Milestone requirement - “month-ahead” probabilistic forecasting

Ensemble forecasting – with ECMWF

Clements, Hurn and Li (2015) state level NEM load forecasting in Queensland

- Day-ahead load forecasting
- Piecewise linear forecasting (9-15, 9-20, 22-26, 22-30 degrees centigrade) like CDD/HDD
- Three year rolling window

Weather data sources

- **European Centre for Medium Range Weather Forecasting (ECMWF)**

The extended-range forecast (HRES) provides an overview of the forecast for the coming 46 days, focusing mainly on the week-to-week changes in the weather. (e.g. windy.com) 0-1104 hours 6-hourly; 0-360 hours at 0.2 degrees; 360-1104 hours at 0.4 degrees. 50 ensemble forecasts.

- **Global Forecasting System (GFS)**

Public domain access – 0.25 degrees resolution, to 384 hours (16 days). (e.g. wxcharts.com)

- **Bureau of Meteorology (BOM)**

APS3 ACCESS-C – 0.0135 degrees (~1.5 km) around cities; hourly to 36 hours

Business Demands for Numerical Weather Prediction

Sector	Sector Needs	Outcome	Output
Energy	"I need to know the amount of solar radiation reaching the ground, so that I can plan the productivity of major solar farms and make high value decisions and ensure that there is enough power available to meet demand. I need timescales of 3 hours to 3 days."	Information on solar radiation up to 3 days for specific locations	**City-scale model** APS: CE3 APS: CE4

Ouija Lite

What is being predicted – usually “Grid Import” interesting for DNSP as opposed to house load and solar generation

In areas of device installation – very high solar penetration

Need different models for different areas

Quantile regression forest generalizing random forest

Forecasting

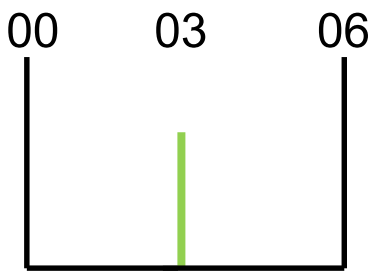
Training on ECMWF data – “**Copernicus Climate Data Store”** data (hourly, worldwide – large data)

Eight variables: Total precipitation, 2 m temperature, surface solar radiation downwards, surface thermal radiation downwards, U10 and V10 (10 m wind speed), total cloud cover, surface pressure

Forecast has 6 hourly resolution – total precipitation, SSRD, STRD cumulative; others point forecasts

Times at 00, 06, 12, and 18 GMT

Develop 24 models using bracketing values e.g. forecast load at 03 GMT using weather values at 00 and 06 GMT



Use weekend Boolean variable and time of year variables

Forecasting

Testing on ECMWF climate forecast – 50 ensembles – 46 day ahead forecast (1104 hours)

ECMWF data downloaded to Azure SFTP server as GRIB v1 (“gridded binary”) and converted to NetCDF for R programs with ECMWF software (“**grib_to_netcdf**”)

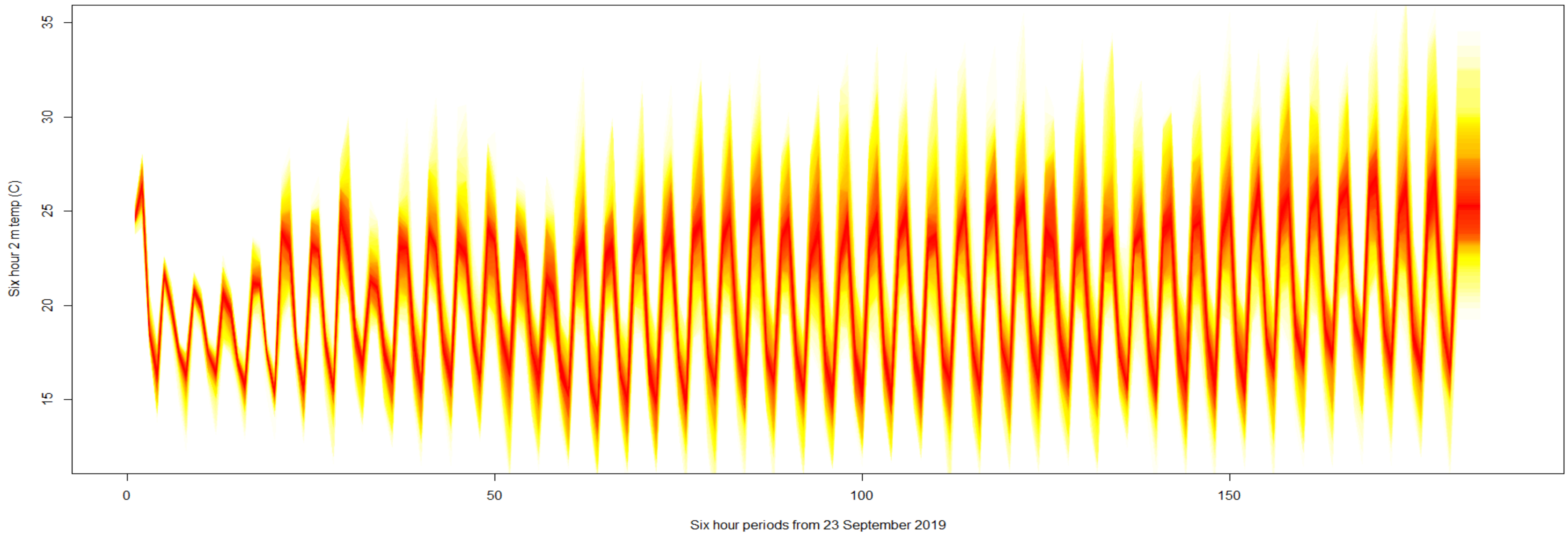
Using quantiles of output as forecast (0 to 100 per cent)

Gaps between training data and testing data until hindcast/reforecast catches up

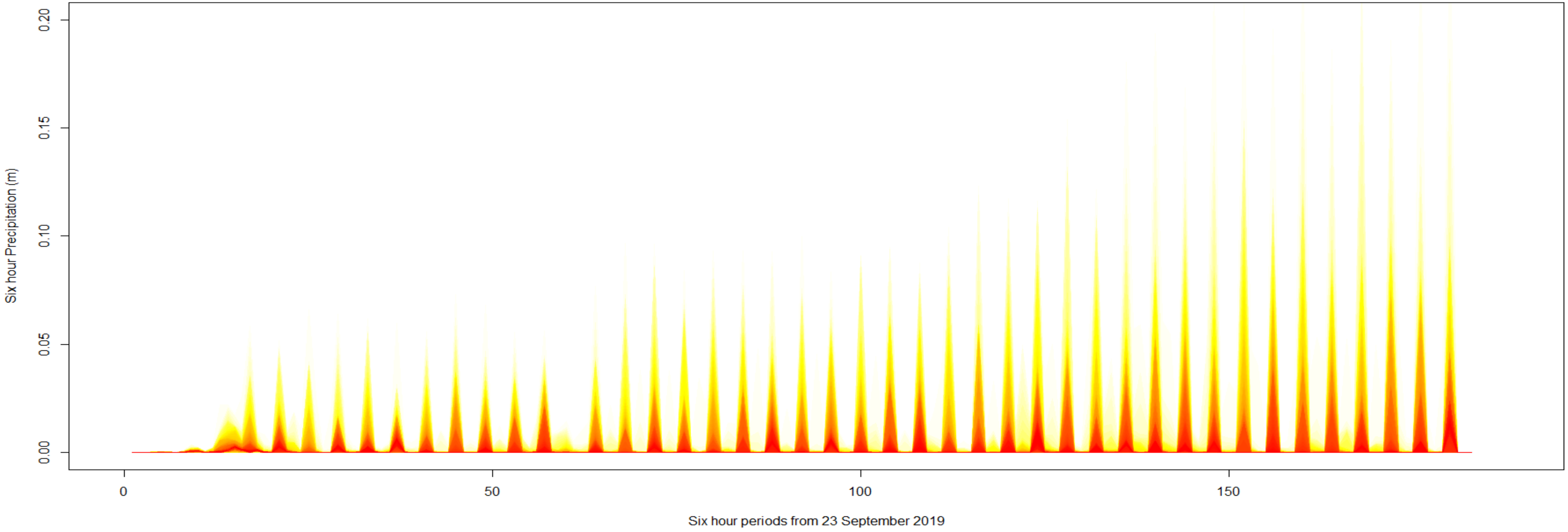
Gaps in demand data availability

50 ensemble forecasts x 101 percentiles = 5050 values for each hour; final percentiles derived from these values

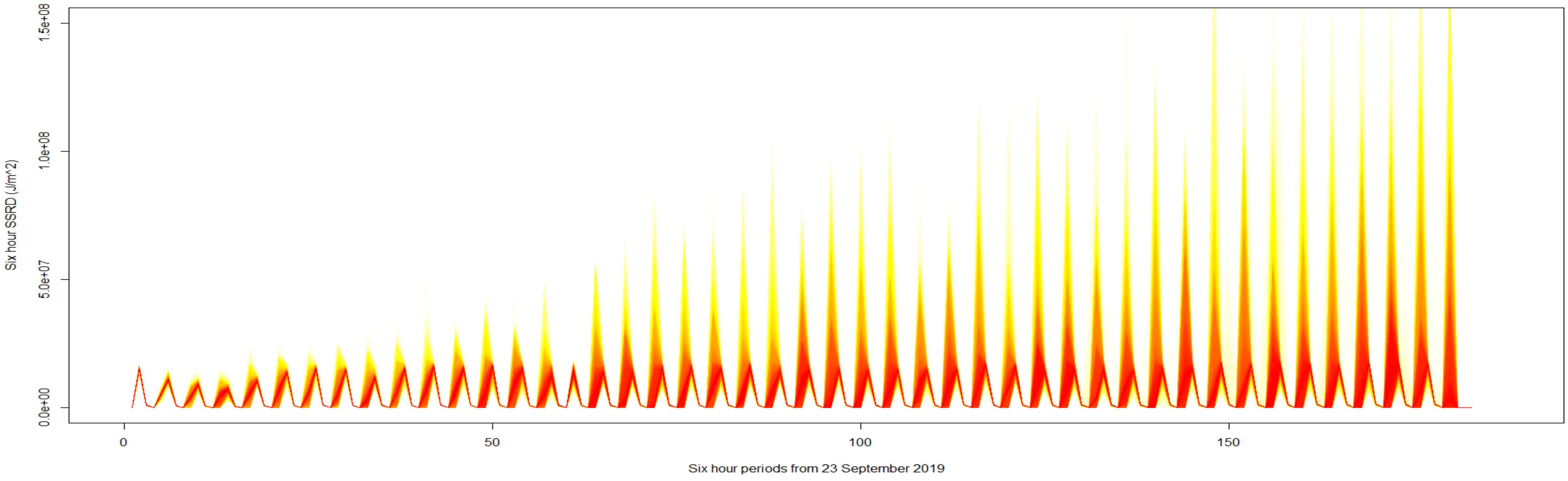
Temperature fan plot



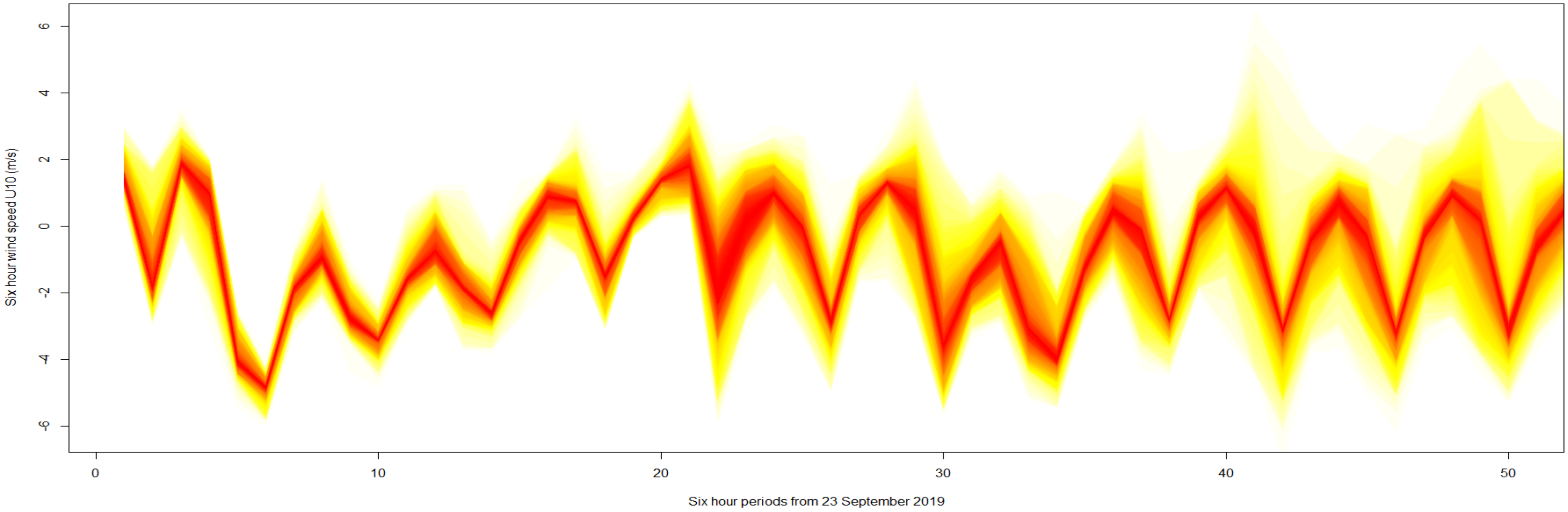
Precipitation fan plot



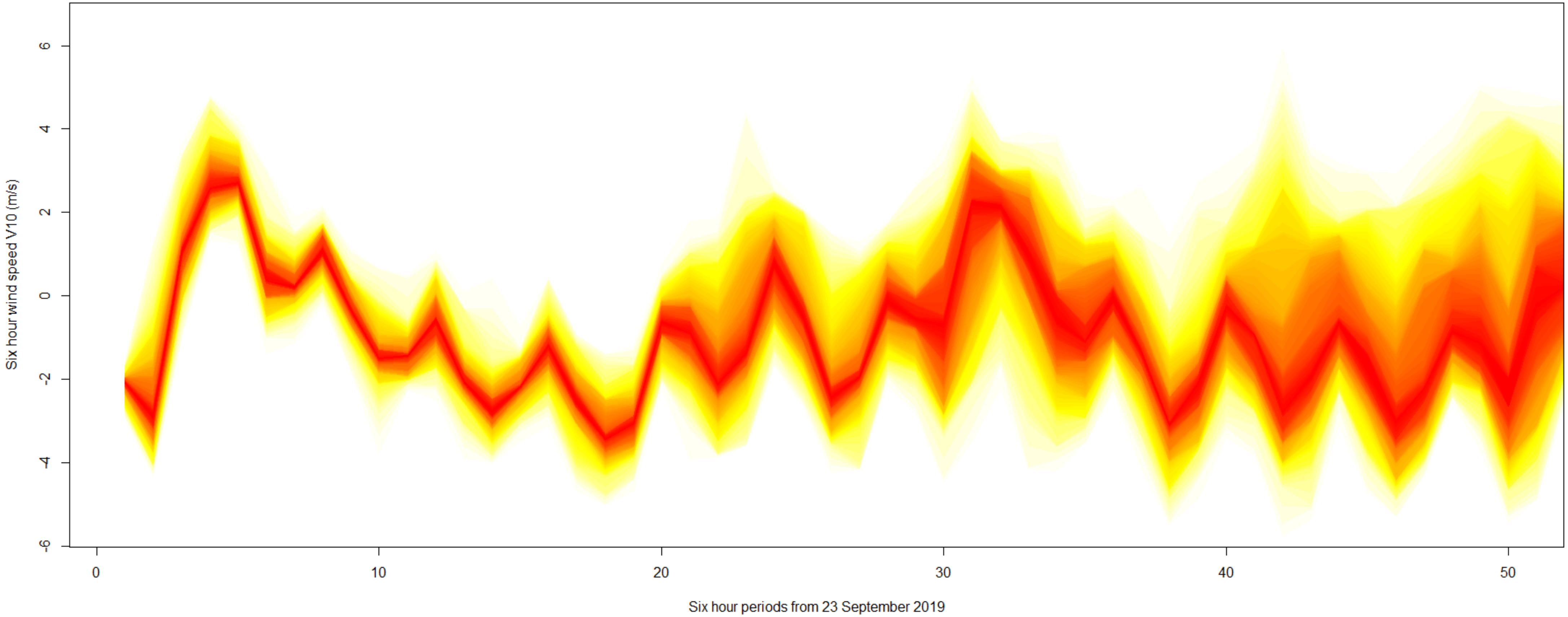
Surface solar radiation downwards fan plot



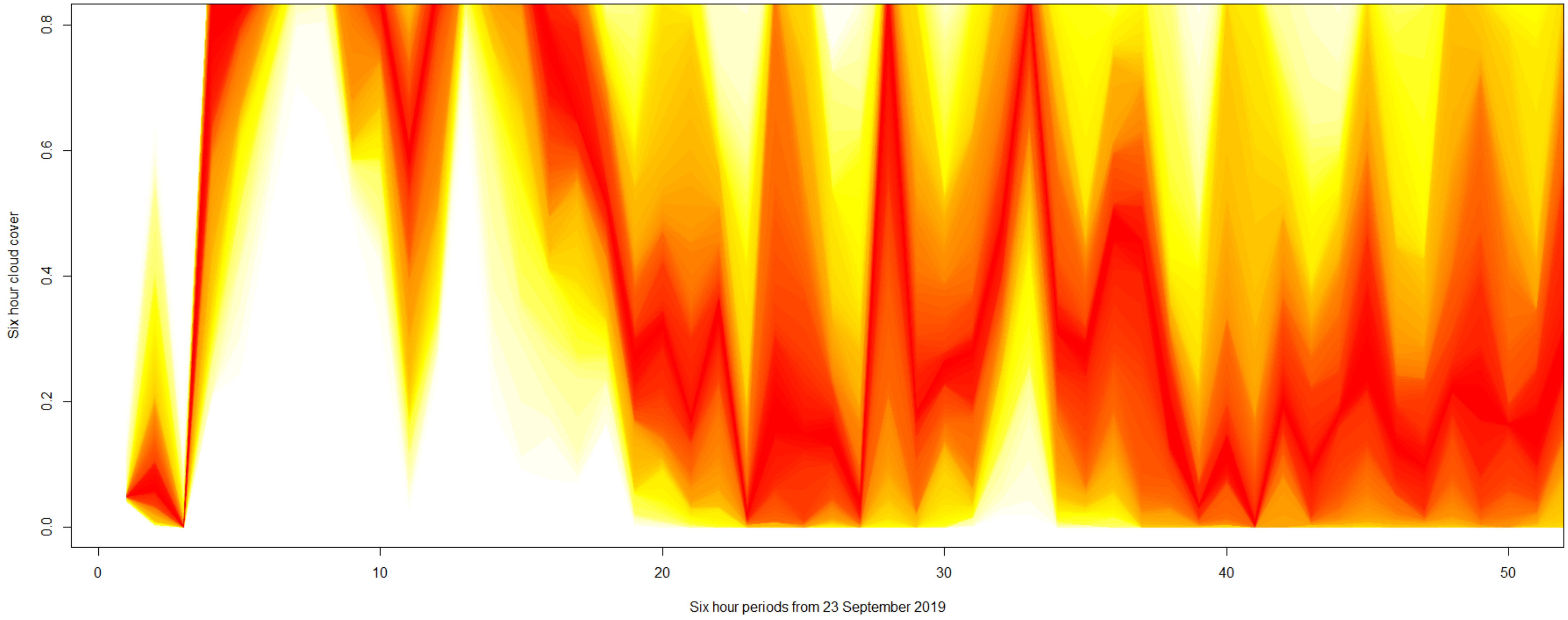
Six hour wind speed U10 (m/s) zonal east-west



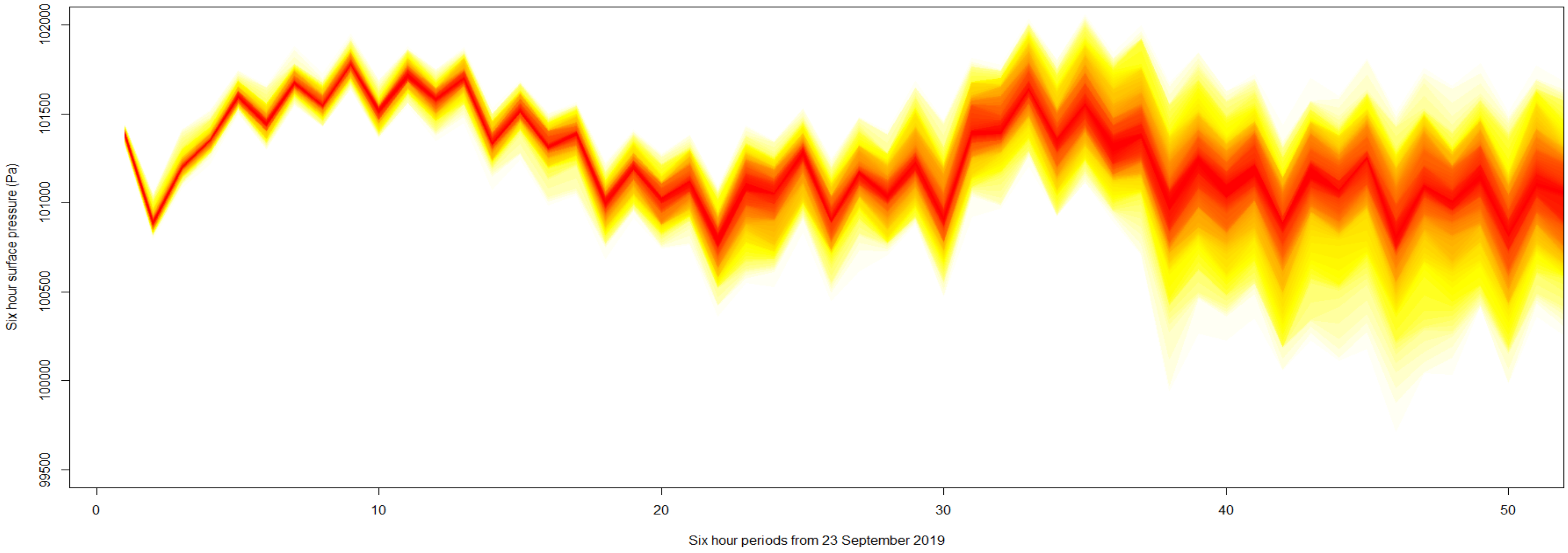
Six hour wind speed V10 (m/s) meridional north-south



Six hour cloud cover



Six hour surface pressure (Pa)



Pinball loss function in Global Energy Forecasting Competition

For each time period over the forecast horizon, the participants needed to provide the 1st, 2nd, ..., 99th percentiles, calling these $q_1, \ldots, q_{99}$, with $q_0 = -\infty$, or the natural lower bound, and $q_{100} = \infty$, or the natural upper bound. The full predictive densities composed by these quantile forecasts were to be evaluated by the quantile score calculated through the pinball loss function.

For a quantile forecast q_a, with $a/100$ as the target quantile, this score L is defined as:

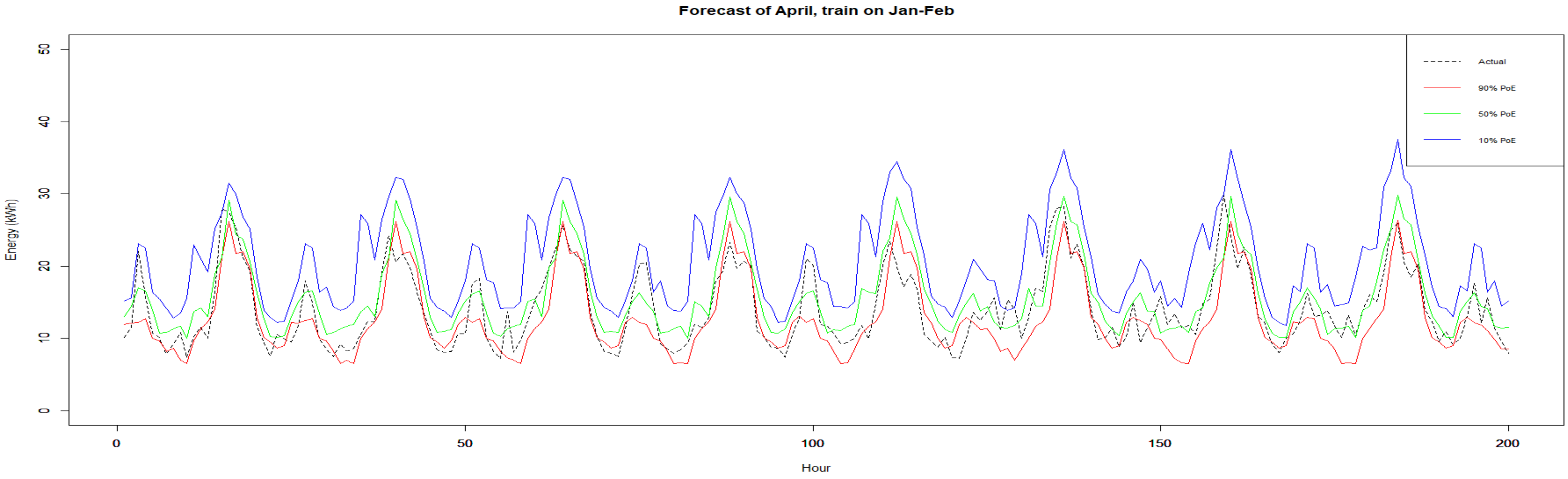
$$L(q_a, y) = \begin{cases} \left(1 - \frac{a}{100}\right)(q_a - y), & \text{if } y < q_a \\ \frac{a}{100}(y - q_a), & \text{if } y \geq q_a, \end{cases}$$

where y is the observation used for forecast evaluation, and $a = 1, 2, \ldots, 99$.

To evaluate the full predictive densities, this score is then averaged over all target quantiles for all time periods over the forecast horizon and for all zones. A lower score indicates a better forecast.

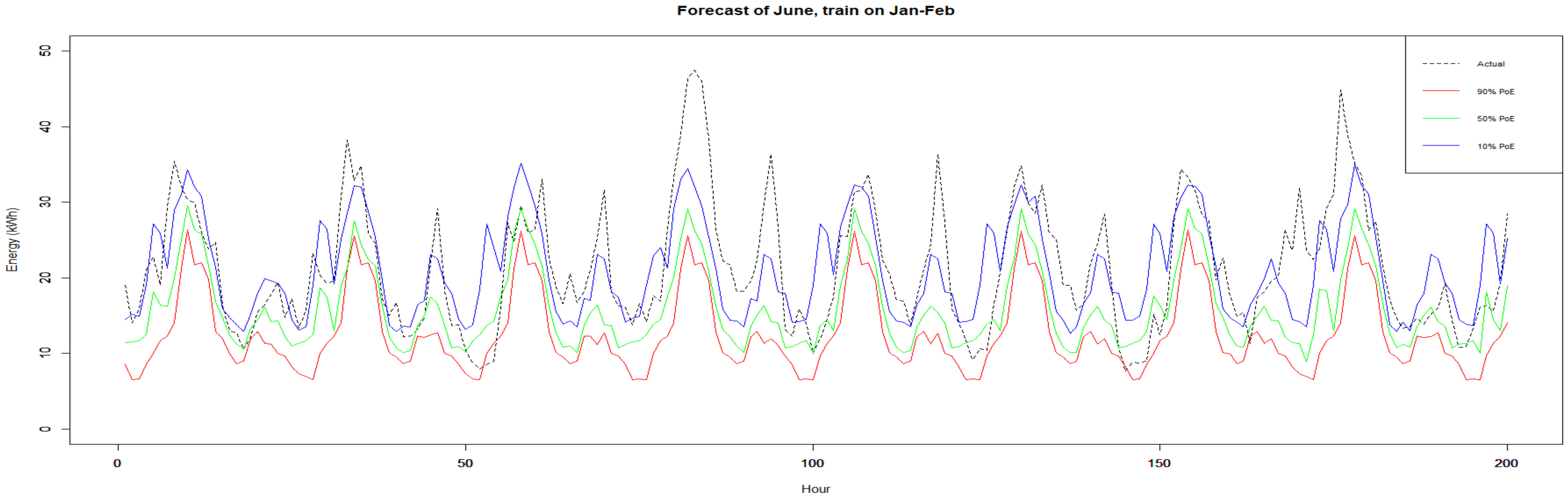
Evaluation – April forecast

Pinball loss function LF = 0.89 using *actual* weather variables

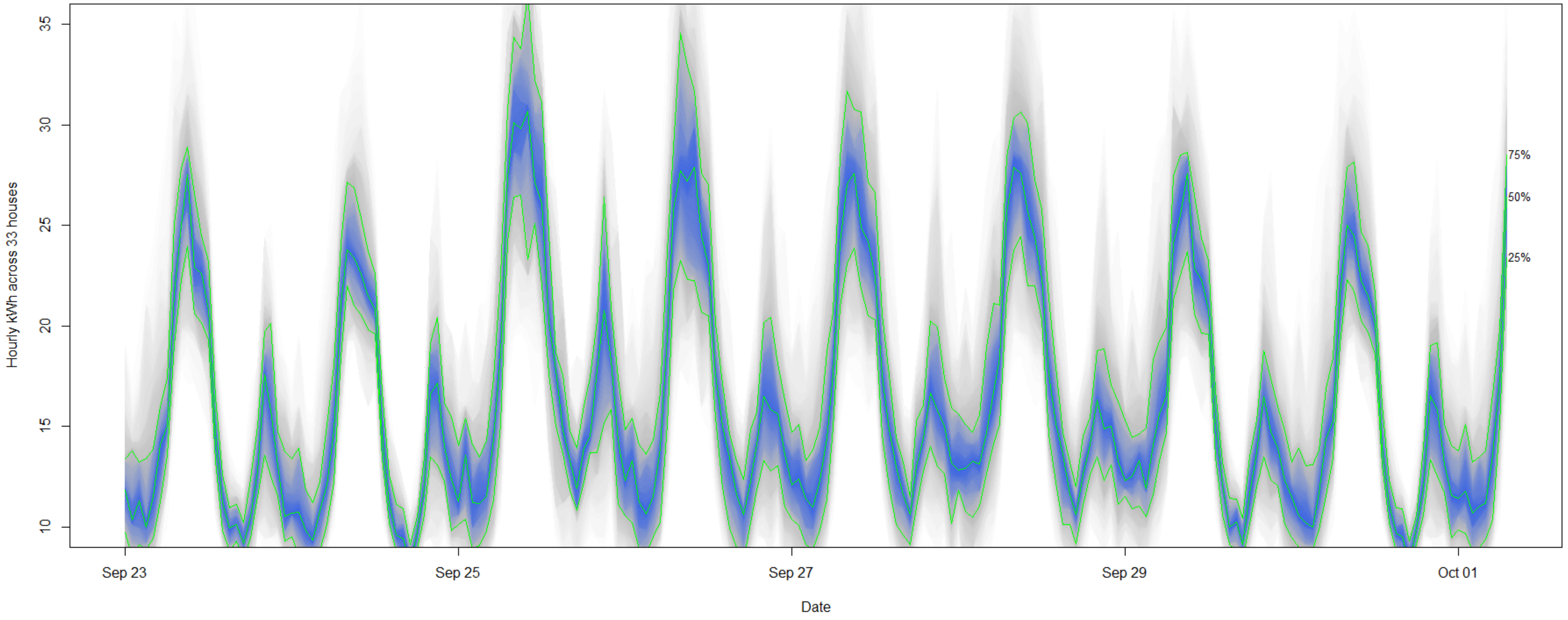


Evaluation – June forecast

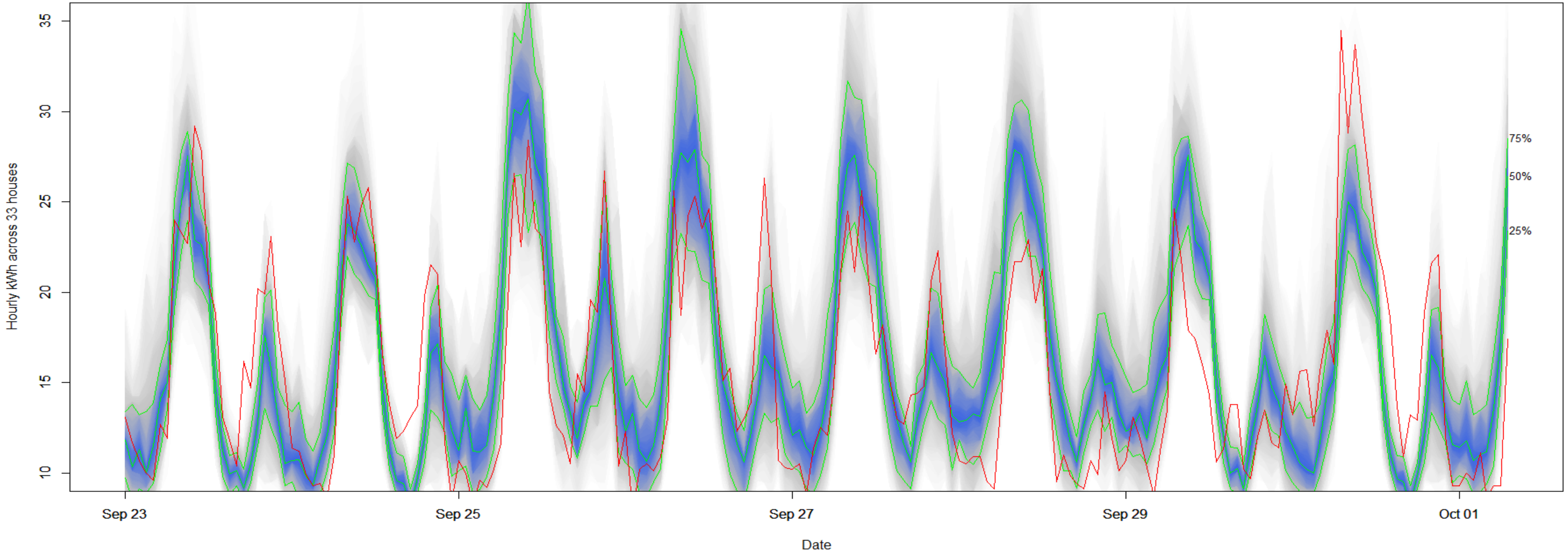
Training on summer months and forecasting on winter. Model assumes lower temperatures mean lower electricity demand. Variable importance rank indicates “t2m” most important, then e.g. SSRD, SP. LF = 2.5



Evaluation – September forecast (train Jan-Jun)



Evaluation – September forecast (train Jan-Jun) actual load superimposed



Limitations

Quantile regression forest can't forecast what hasn't been seen before i.e.

- Training on summer, winter results aren't very good
- Hourly models' minimum and maximum are limited to what has been seen before

Implication – best to have a year's worth of training data, as for Clements et al NEM model; but can train autumn and forecast spring

Hourly forecast, but resolution of input data (with diversity) is two-hourly with six-hourly weather forecast

Need to be careful that input weather data uses same model as forecast data – can differ

Further work/research

- Practical considerations – Python has better integration with Azure than R
- Legal considerations – anonymization and protection of data
- Data considerations – need more data; an accurate “ground truth” information for phase ID and a consistent source of data for aggregate load on a feeder
- Consultation with EQ and DNSPs – what aggregation level is useful, operationalizable; prediction intervals
- Non-intrusive load monitoring (NILM) – accuracy and completeness
- State monitoring of networks
- Visualizations – e.g. different kinds of customer, projected to two dimensions
- Broken neutral detection via impedance
- Better understanding of metering and how it relates to billing (parallel with Metering and Settlement)
- Forecasting system interruptions due to severe weather conditions
- Forecasting five-minute price data in NEM
- Publishing papers

Thank you

Dr Richard Bean | Research Fellow
School of ITEE
r.bean1@uq.edu.au

facebook.com/uniofqld

Instagram.com/uniofqld

linkedin.com/in/richardbean