THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# Phase identification and load forecasting with home energy data

New Insights for Network Operations Using Big Data


## Richard Bean

Centre for Energy Data Innovation (CEDI), The University of Queensland


19 March 2020, EQ, Newstead

Energy Queensland

REDBACK TECHNOLOGIES

SPRINGFIELD CITY GROUP

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

Microsoft

ADVANCE QUEENSLAND

# Centre for Energy Data Innovation

- Research enterprise created in partnership between The University of Queensland and Redback Technologies.
  - AI and Data Analytics
    - Demand Forecasting
    - Non-intrusive Load Monitoring (NILM)
    - Fault detection
    - Safety assessment
    - Consumer feedback
  - Advanced Visualisations
    - Energy dashboards
    - Augmented reality
  - Privacy and Security
    - Smart grid security risks
    - Consent mechanisms for sharing data

# 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

# A word on data

| | High level (Monthly data) | Mid level (30 minute) | Detailed level (1 minute or less) |
|---|---|---|---|
| | Broad management | Data driven | Big data driven |
| Customers | Market management | Customer Management | Appliance management |
| Networks | NEM Management | Substation Management | LV Feeder Management |
| Other | Policy setting | Investment decisions | Real time efficiency |
| | | Consumption (kWh) | Voltage, Current, Impedance, power factor |

# Redback Ouija Lite devices

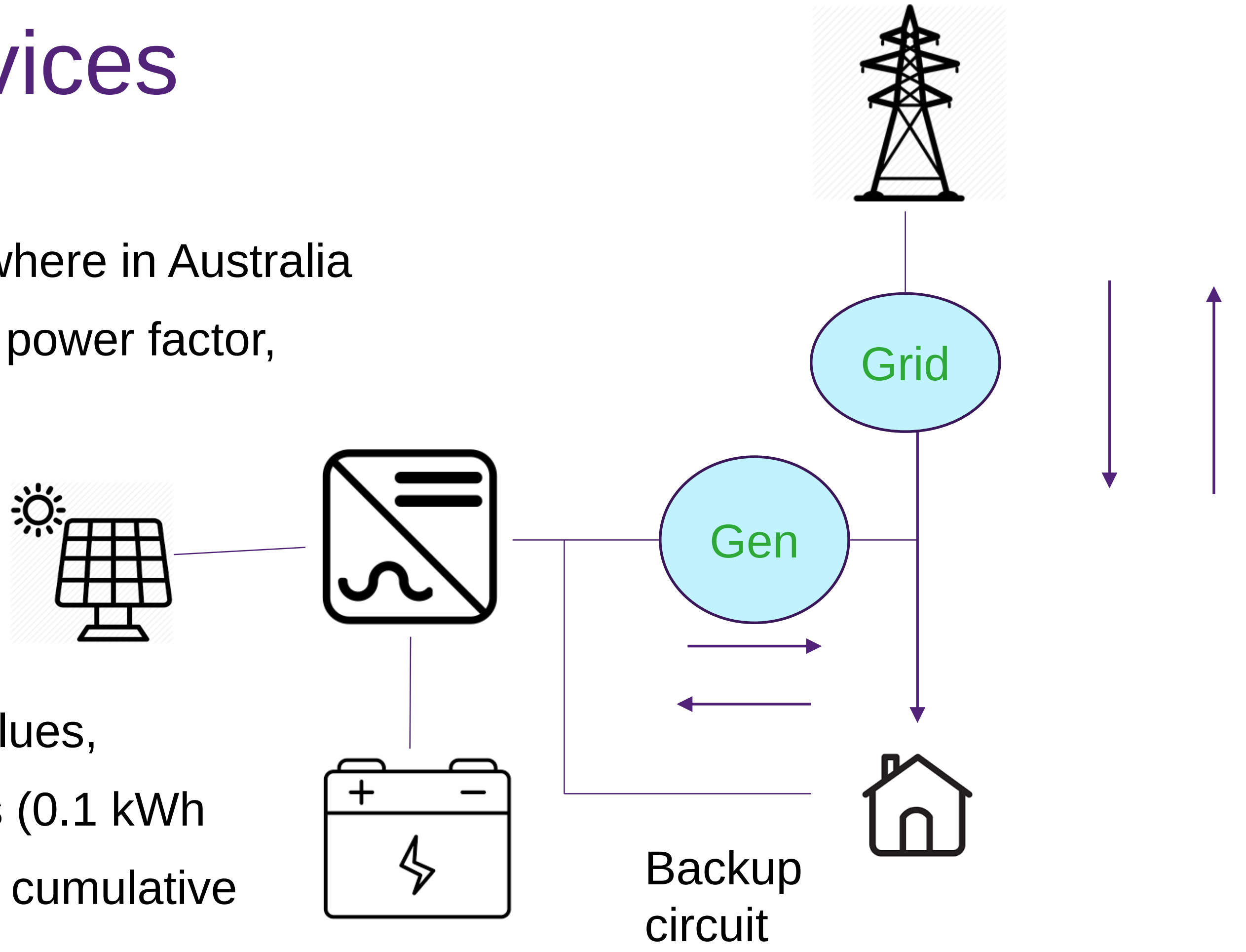
*What devices record*

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.

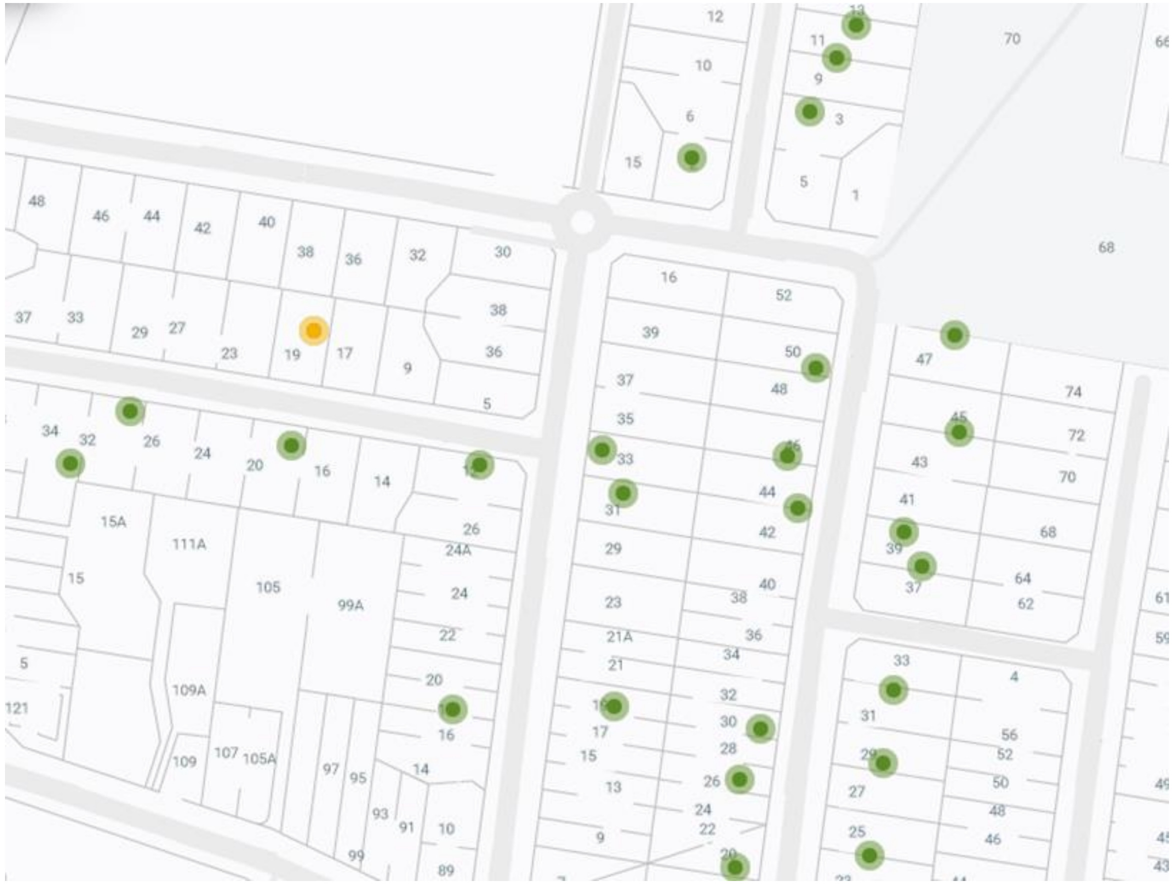
One minute resolution for instantaneous values, cumulative values updated every two hours (0.1 kWh resolution). But diversity in summing house cumulative values.



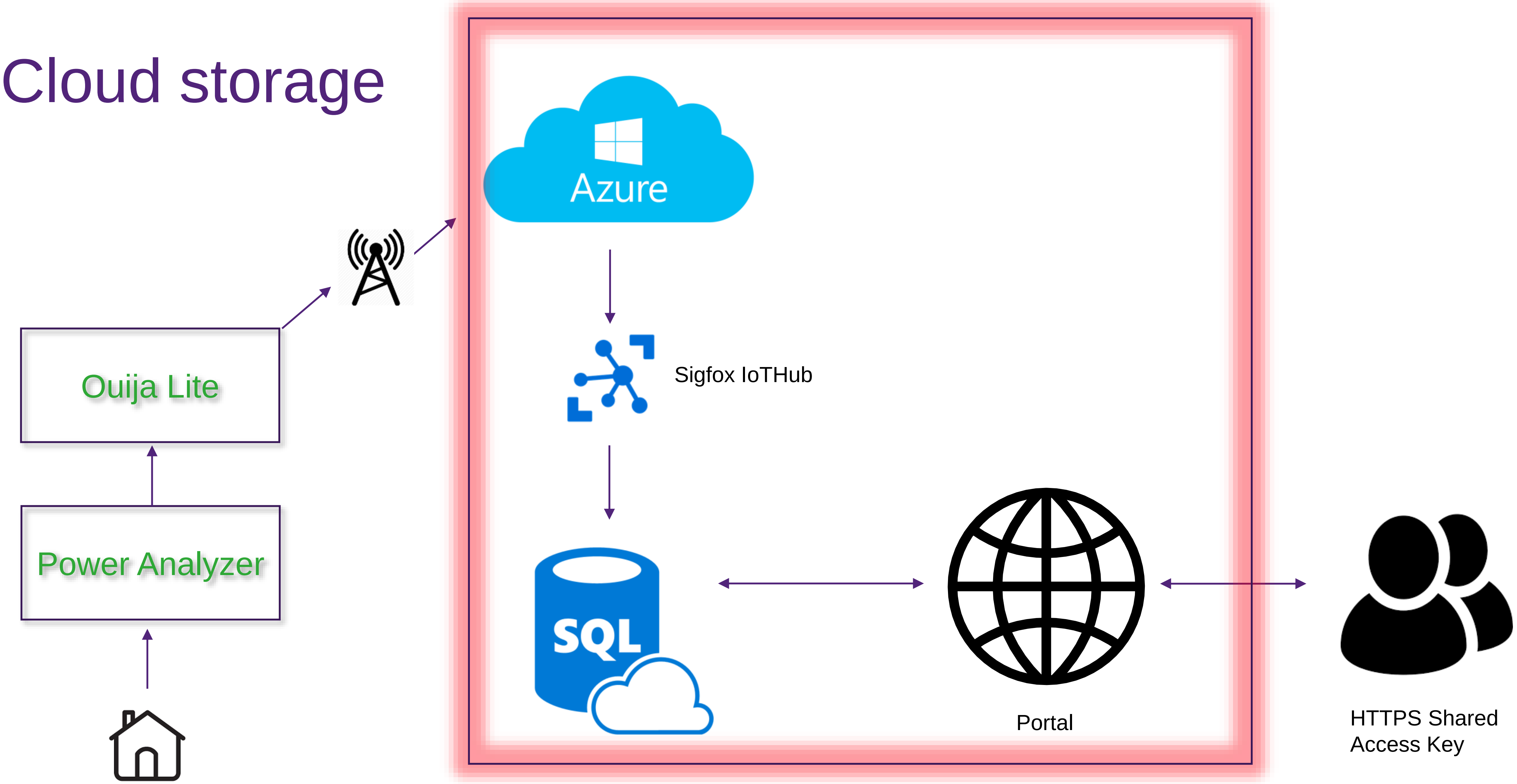
# Exploring the physics of LV feeders in real time

Power Data

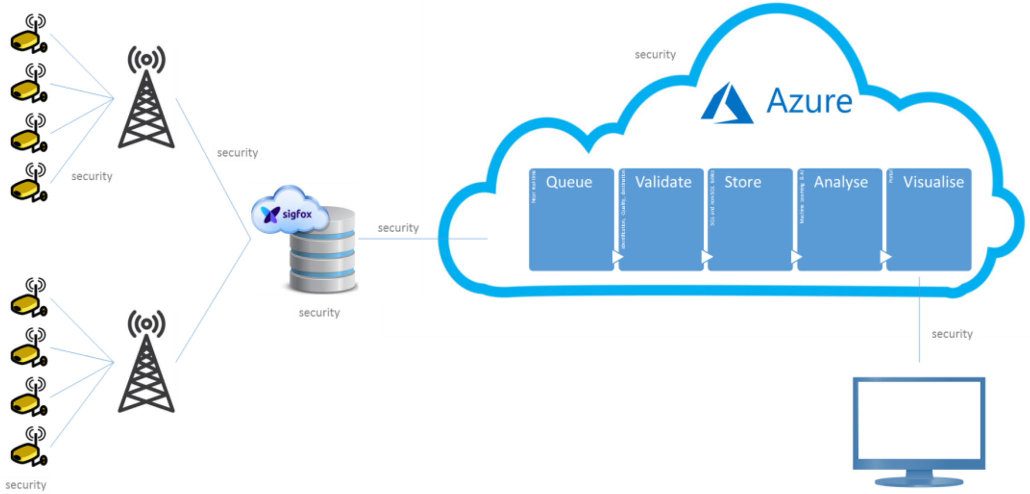
- Current (amps) (import/export)
- Voltage (volts) (import/export)
- Power (watts)
- Impedance (ohms)
- Power Factor



# Cloud storage



# Knowledge




Expertise
- Network Engineers (EQL)
- Network ICT (EQL)
- Communications Engineers (Sigfox)
- Software Engineers (Redback)
- Hardware Engineers (Redback)

Research
- Power Engineering
  - Future Grid, Solar PV and Batteries, VPP
- Data Science
  - Platform design and scaling, Data storage and compression, Machine Learning and Artificial Intelligence, Security
- Interaction Design
  - Human-computer interface, CX, Service design, visualisation, AR.

# Progress to date

- First 100 devices
  - Understanding the data
  - Understanding the problems (data collection , storage, analysis)
  - Exploring the anomalies
- Opening the Communications channel
  - Ironing out the problems
- Learning about installs
  - Finding room on the panel
  - Accessing meters
  - Asbestos
  - We moved to providing a separate enclosure

- 1st Generation platform
  - User interface designs
  - Scale and pressure testing with Bots
- 2nd generation design
  - Integrating alarms
    - Voltage
    - Broken neutrals
  - User interface customisation
  - API data retrieval

# Progress to date

- Next 500 devices
  - Revisiting installation difficulties
  - Expanded data sets
    - 5,000,000 data sets per week
  - Stronger algorithms
  - Business case for step change roll out

- Visualisation
  - Converting data into understanding
  - Workshopping with networks
  - Prototyping substation health indices
  - Exploring AR for feeder interrogation
- Integrating other data
  - Weather data
- Exploring other comms channels
  - 3G, 4G

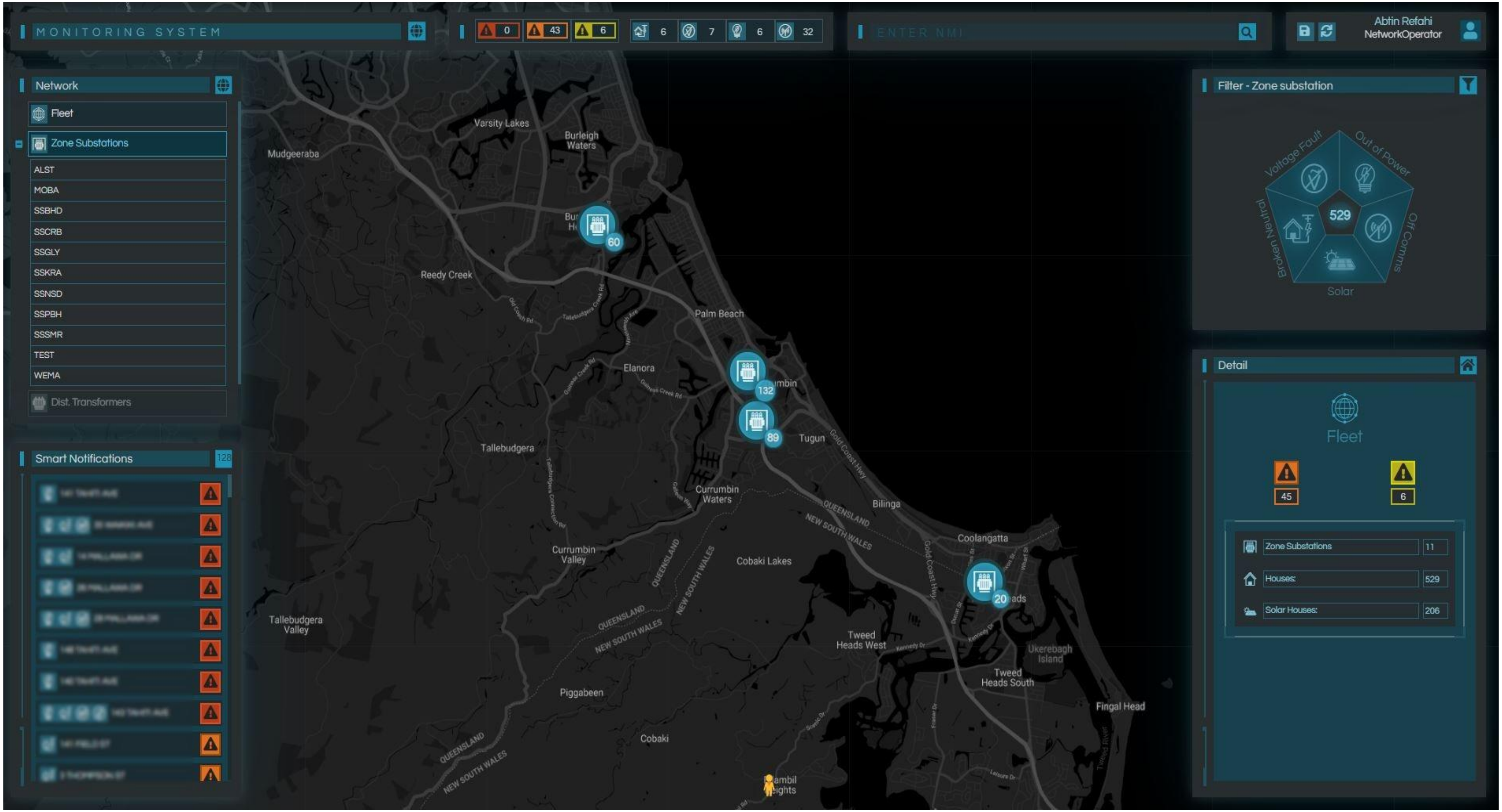
# Next steps

- Data step up
  - 20,000 devices
  - 1 million data sets per hour

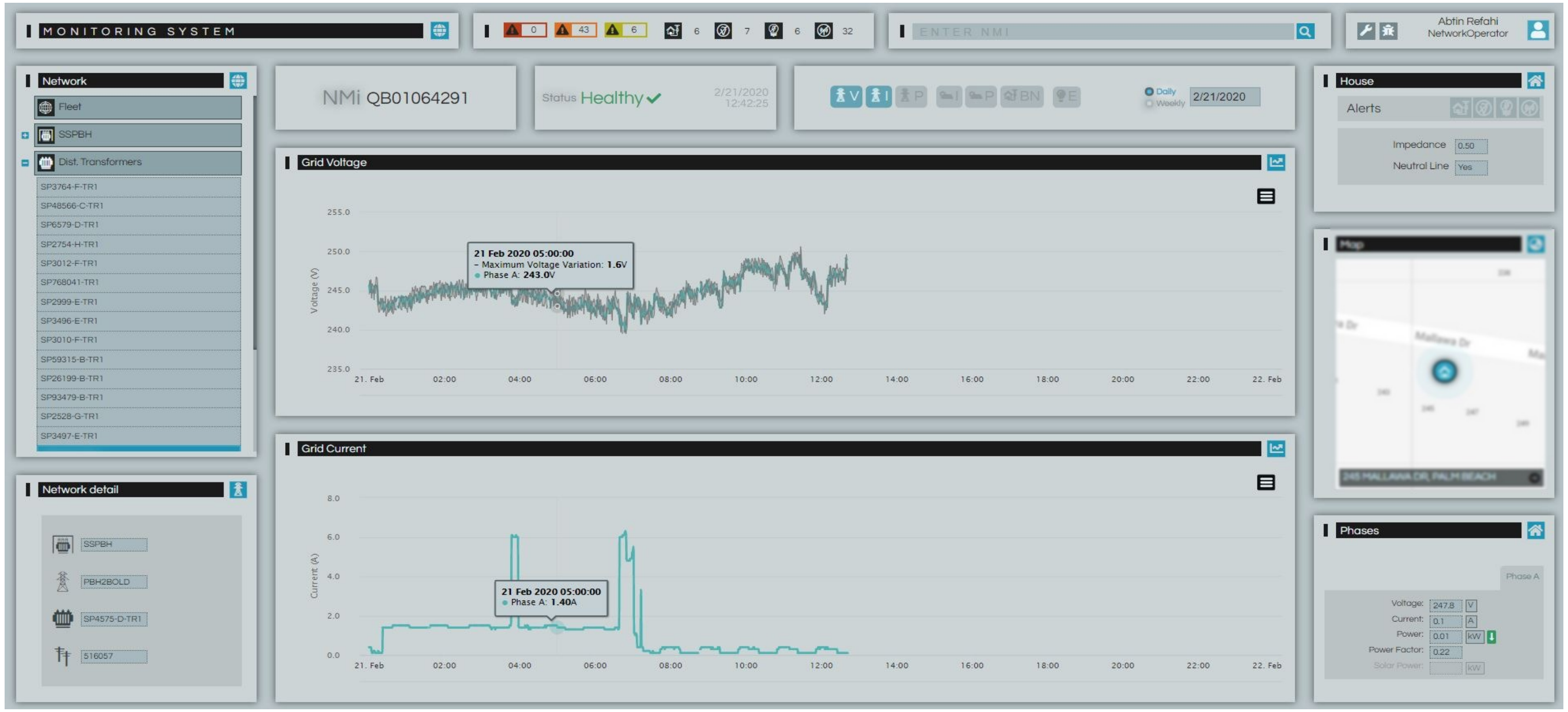
  - Analytics step up
    - Additional broken neutral algorithms
    - Condition monitoring through signal deterioration
    - High impedance faults
    - Supply to life support customers
    - Real time emergency response

- Receiving data from multiple sources
  - Redback Devices
  - Smart Meters
  - Network data
  - Inverters
  - Other?
- New comms channels
  - 3G, 4G

# Dashboard view



# Dashboard view




Power Data

- Current (amps) (import/export)
- Voltage (volts) (import/export)
- Power (watts)
- Impedance (ohms)
- Power Factor

# 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

# 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

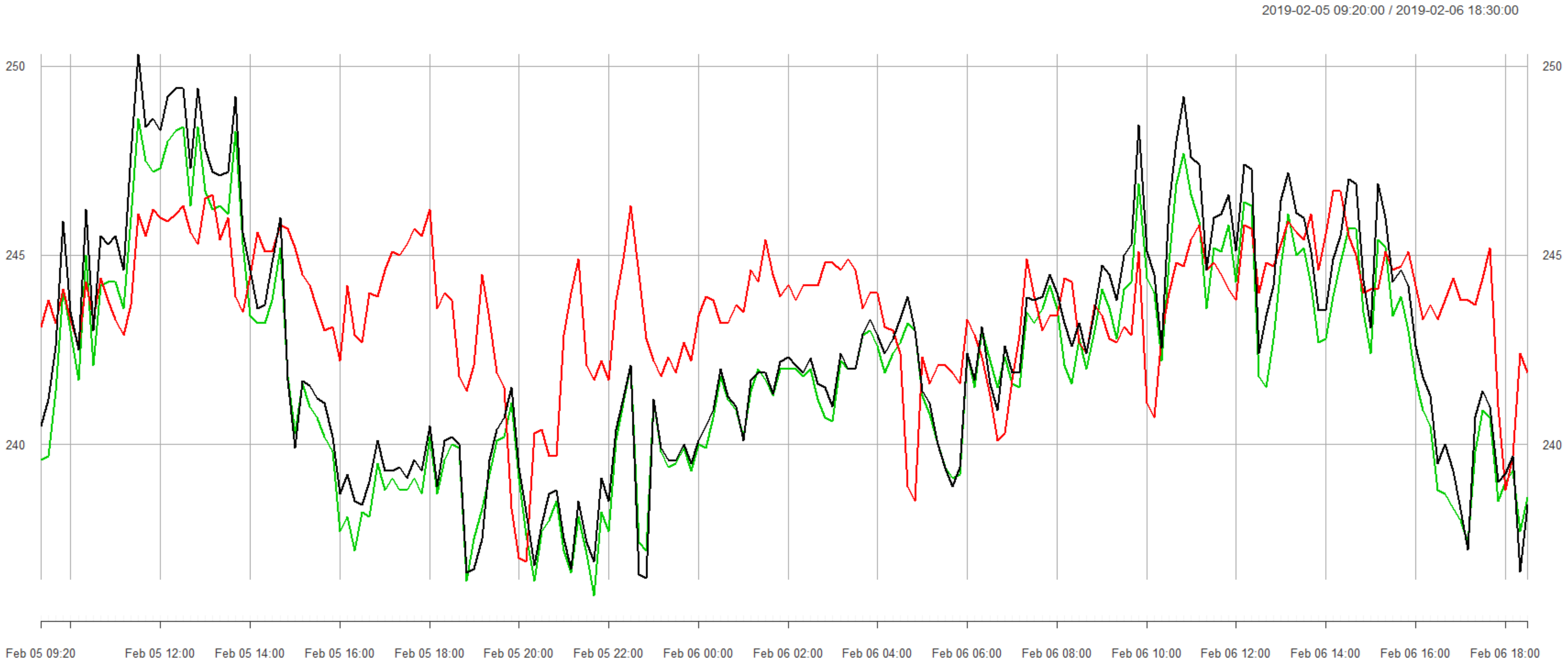
Limited data – good data hard to come by – P & W had 74 houses

Legal obstacles – privacy considerations – addresses/NMIs as identifiers

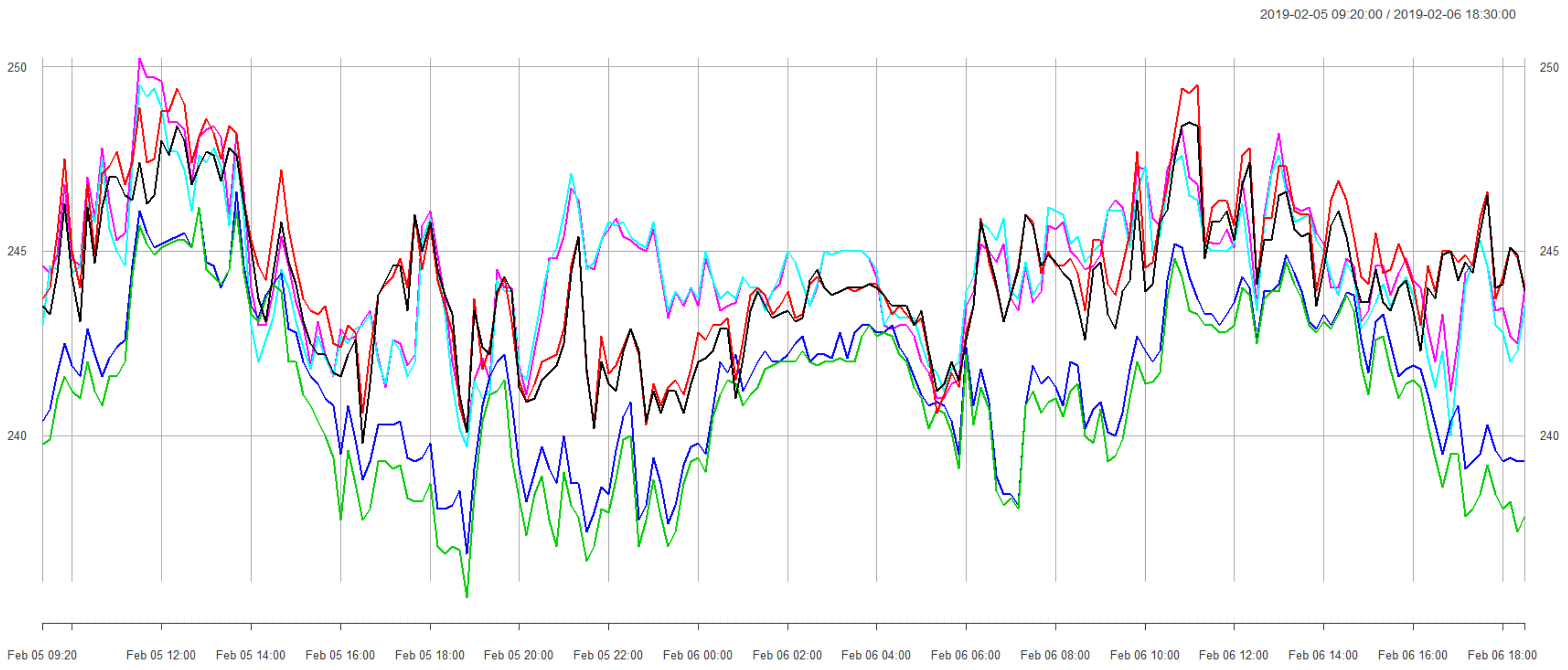
Unsupervised clustering – partitioning around medoids (PAM – Kaufman and Rousseuw 1987)

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

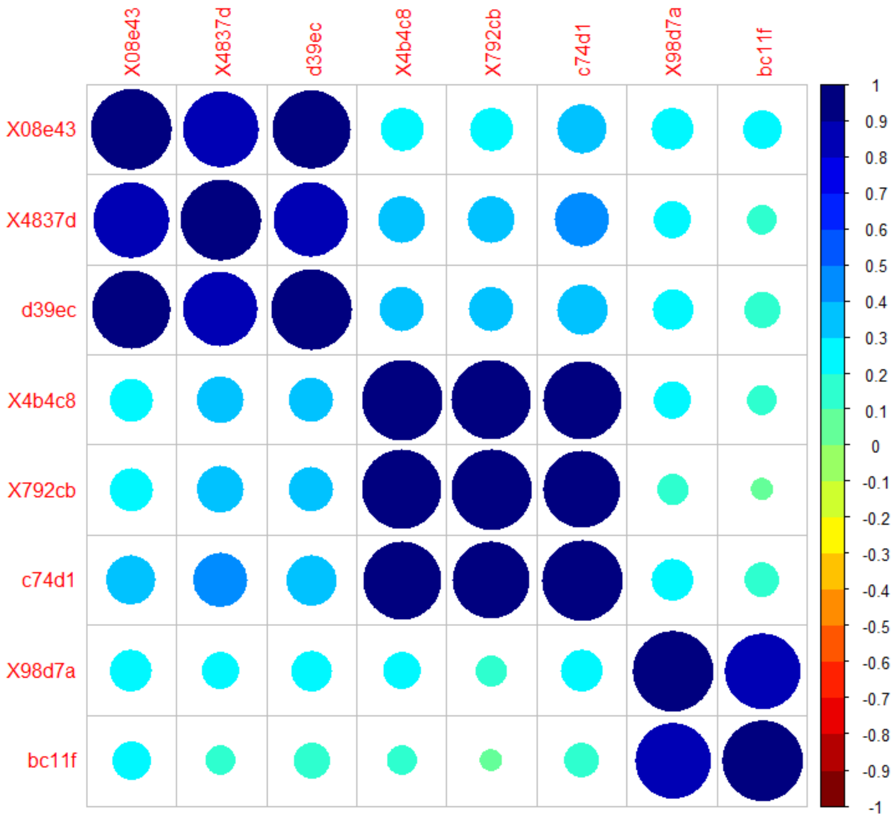
# Phase ID time series – 10-minute aggregation



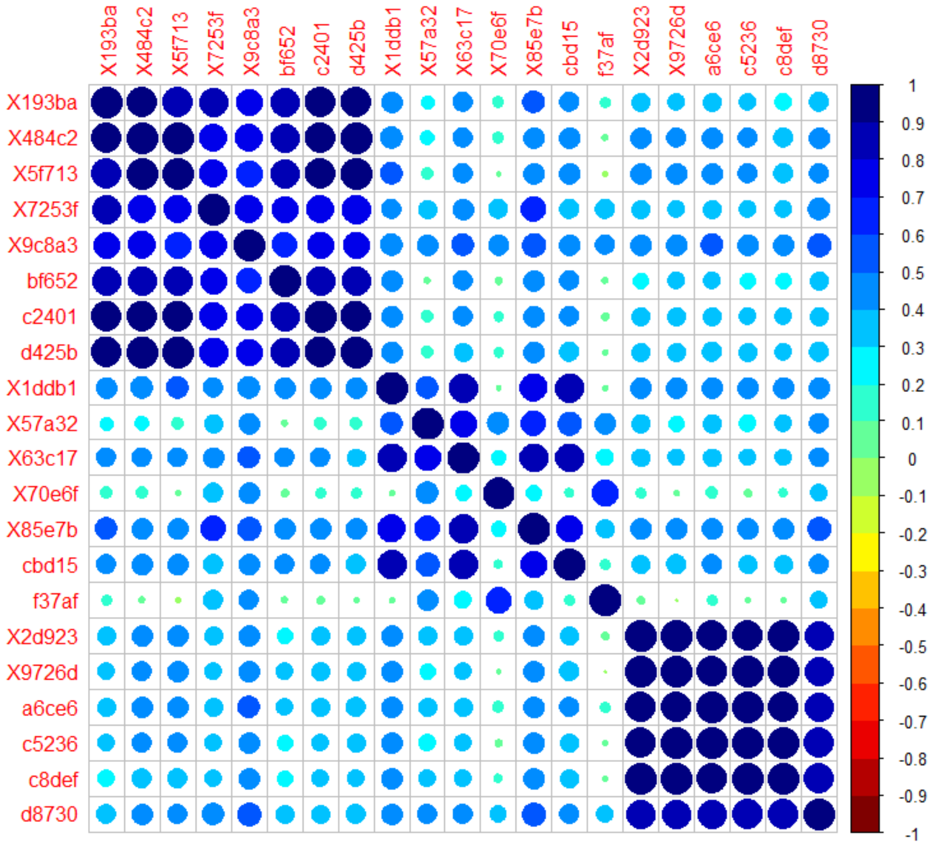
# Phase ID time series – 10-minute aggregation



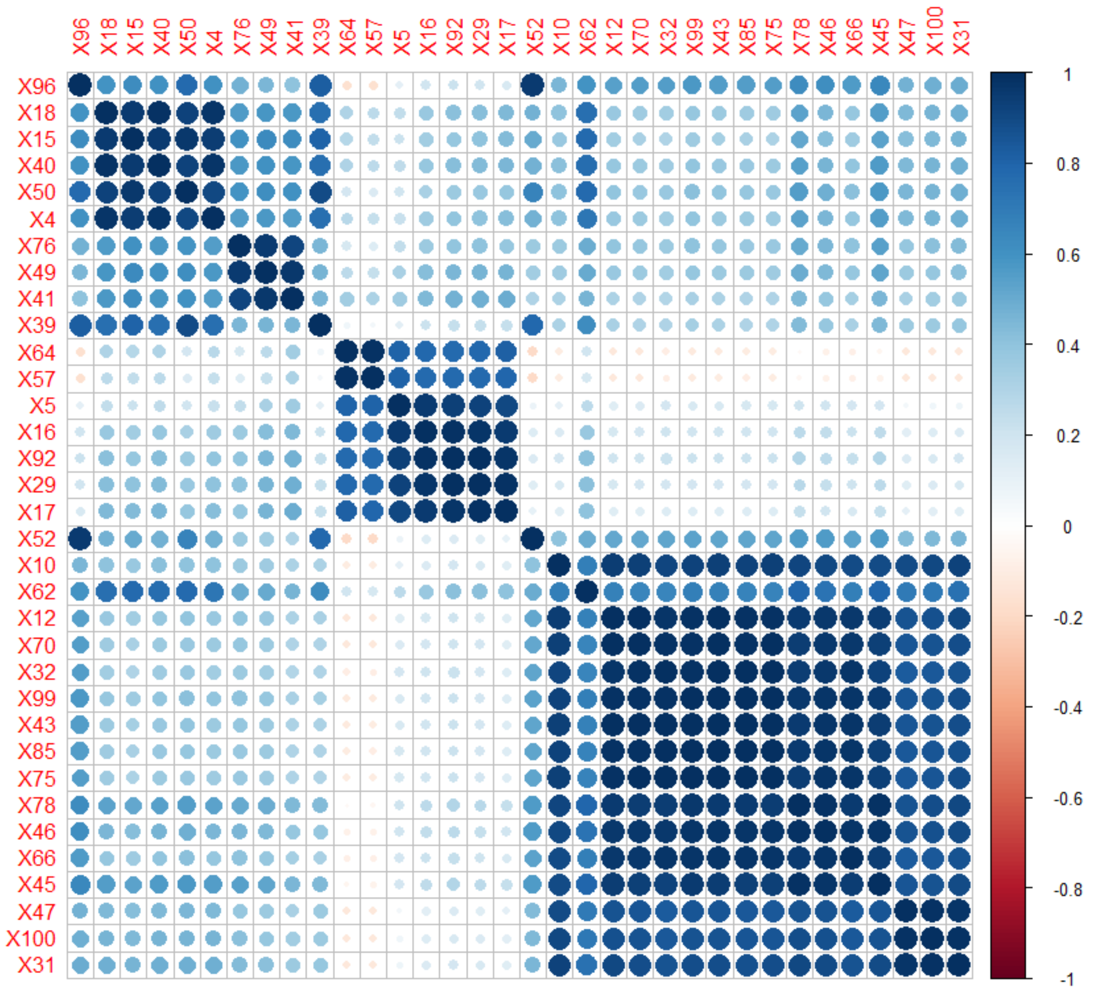
# Phase identification – correlation plot - eight houses

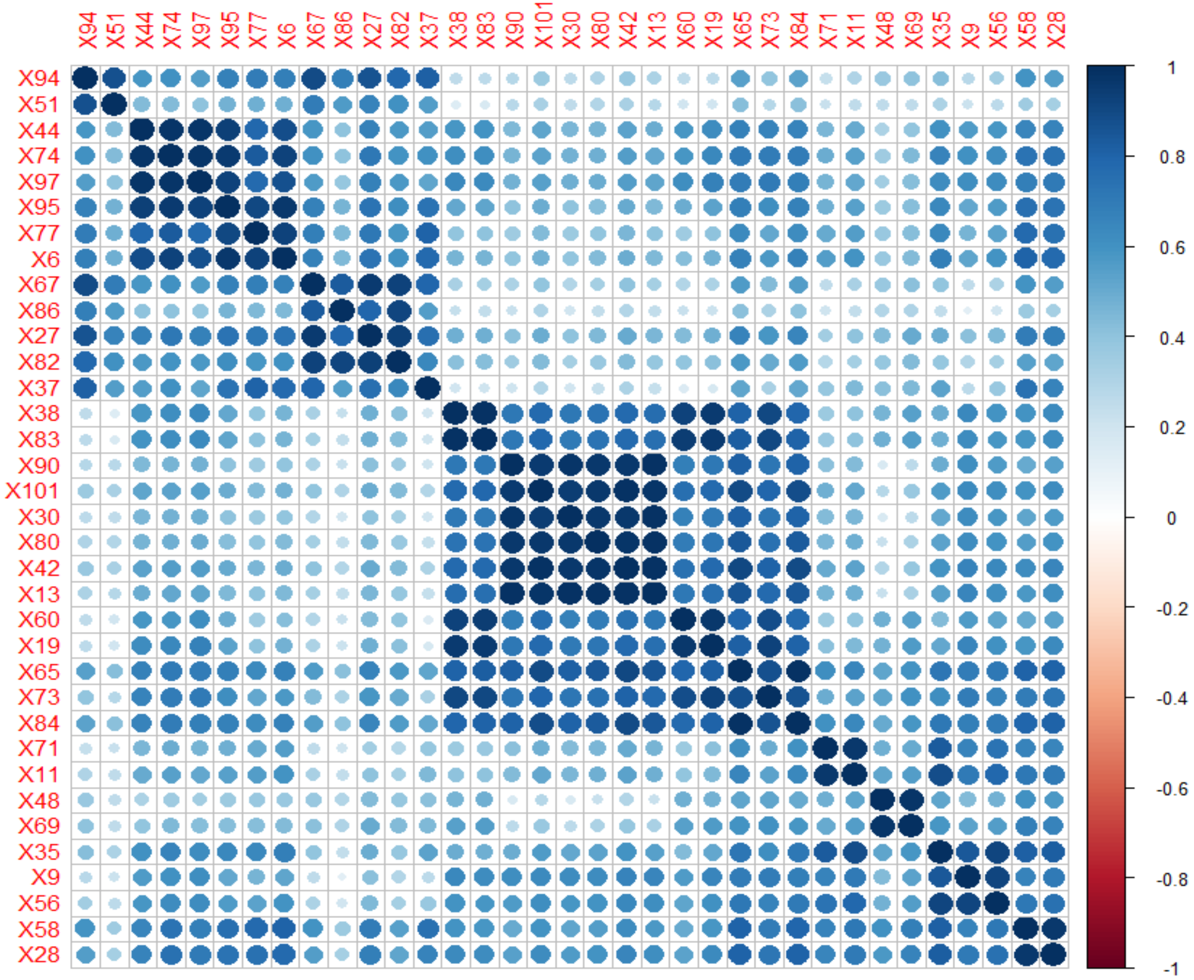


# Phase identification – 21 houses

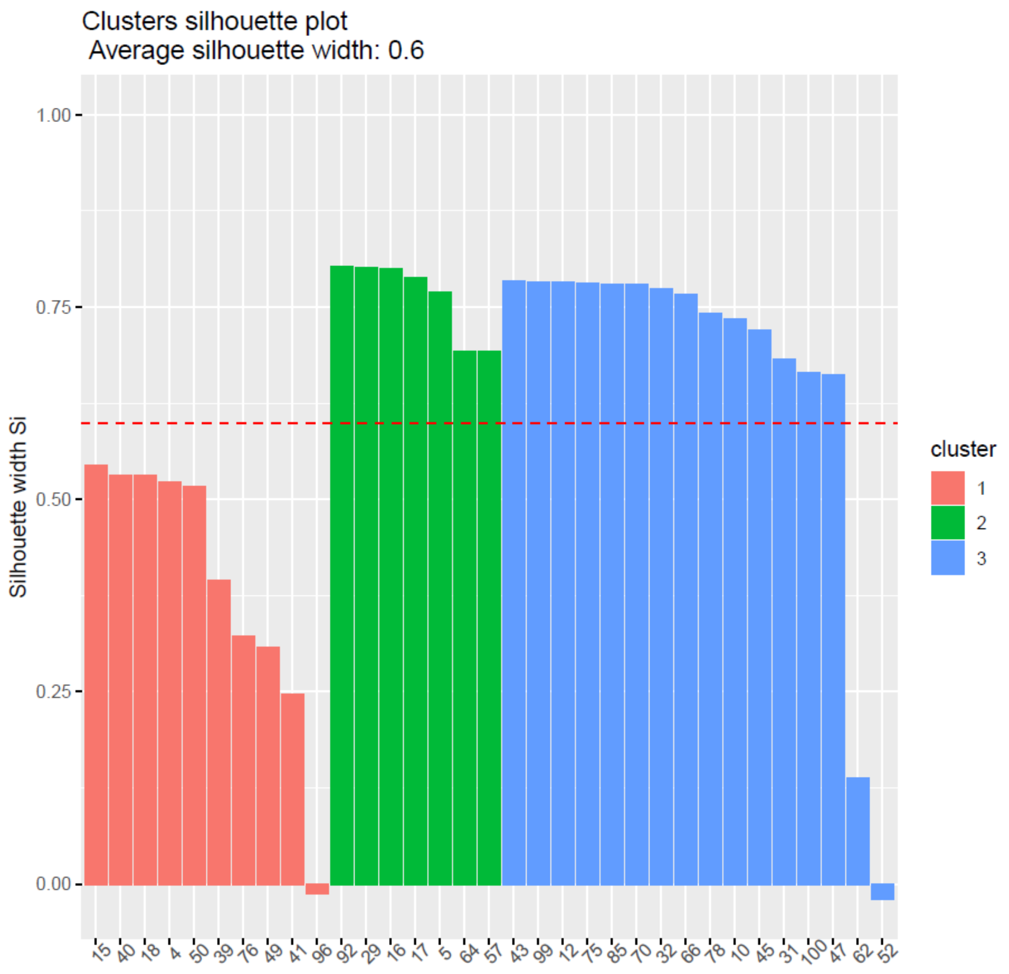


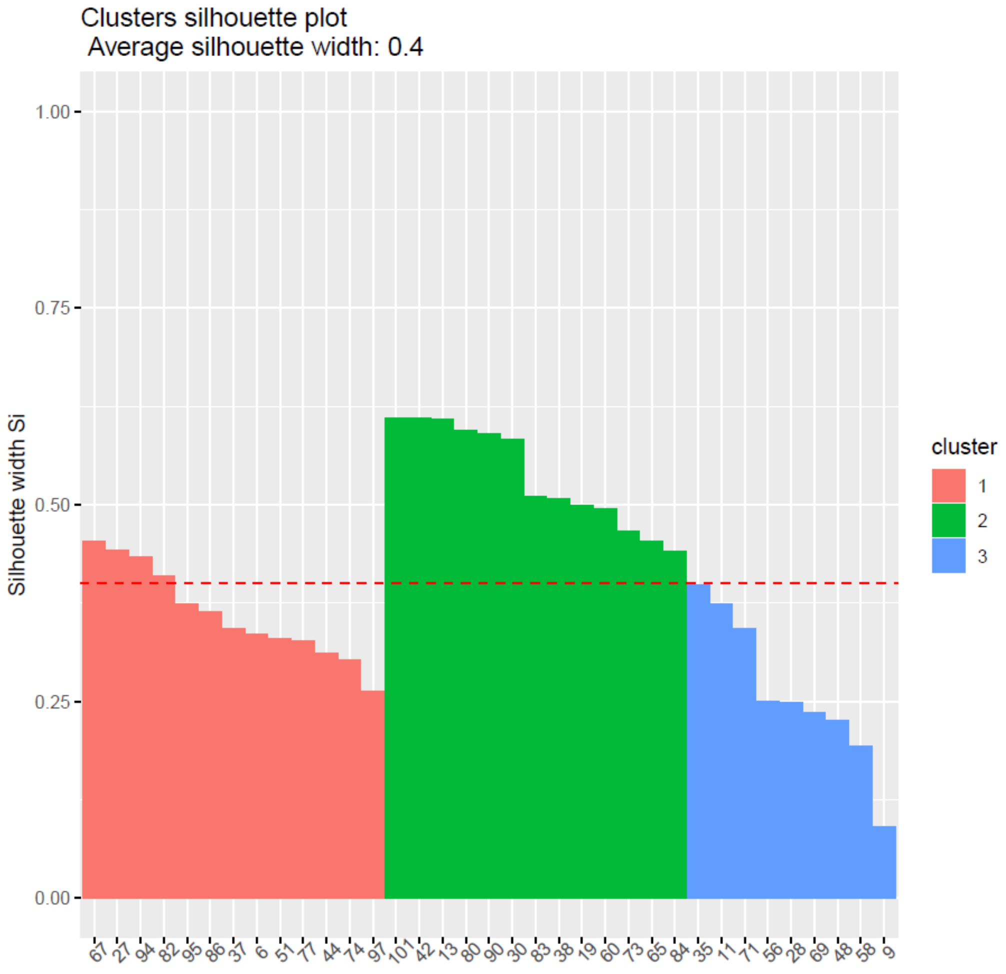
# Phase identification – Mitchelton houses



# Phase identification – silhouette plot



# Load and Solar Forecasting

- Global Energy Forecasting Competition (GEFcom)
- Milestones of Centre – short and medium term forecasting
- Parsimonious modelling
- Best performing models versus interpretable models
  - “Explainability” and accuracy
  - High explainability
    - ❑ Linear regression
    - ❑ Additive models
  - High accuracy
    - ❑ Random forests / Extreme gradient boosting as “black boxes”
    - ❑ Visualizations and “understandability”
- Operationalizing models (C#, Java, Python, R) – integration with Azure vs Netflix, Kaggle winners
- CPU and memory requirements of building and executing models

# Load and Solar Forecasting

- Using weather variables as predictors
  - European Centre for Medium-range Weather Forecasting (ECMWF)
  - Bureau of Meteorology (BOM)
  - Global Forecasting System (GFS)
- 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
- Prediction of hourly aggregate load at various spatial aggregation levels; short term (e.g. hour ahead) to medium term (e.g. 46 day ahead forecast)
- Probabilistic rather than point forecast
- Originally aimed at battery scheduling for inverters – now for benefit of utilities

# 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**<br>APS: CE3<br>APS: CE4 |

# 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

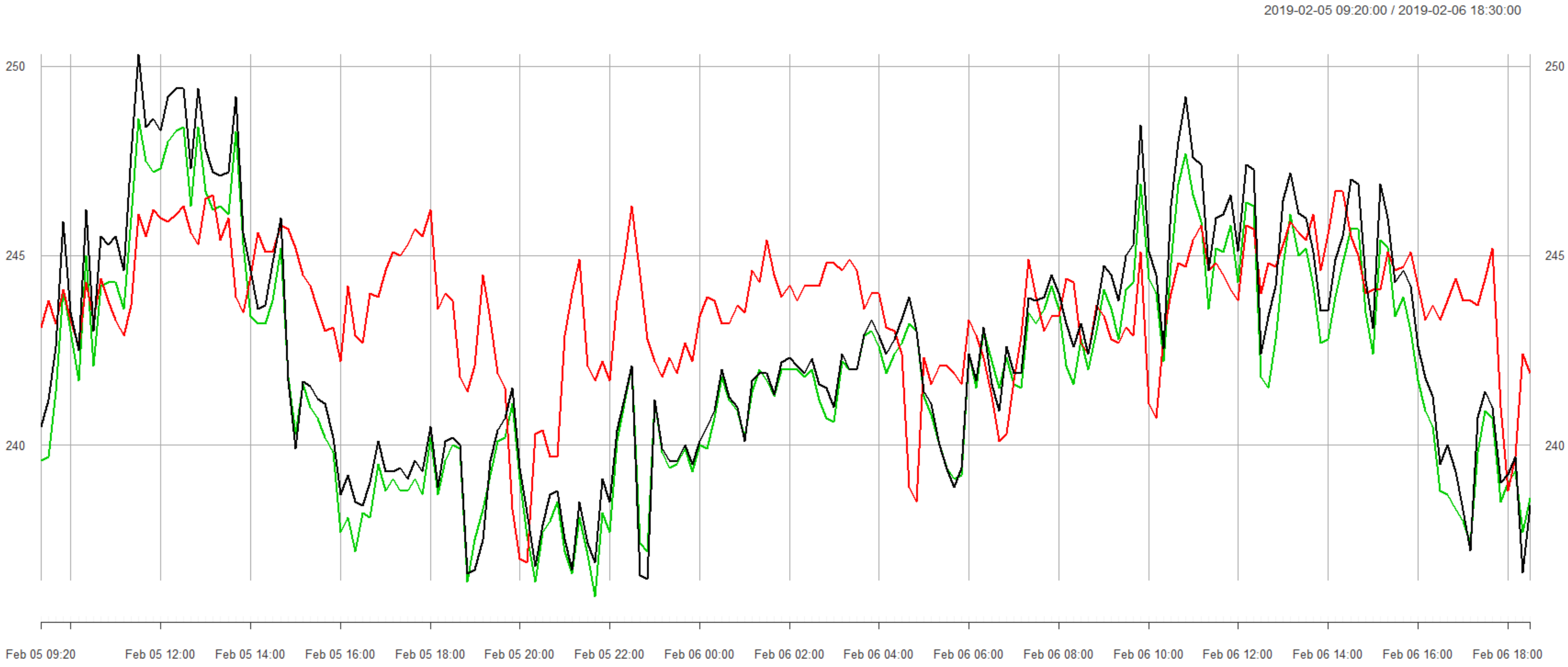
Limited data – good data hard to come by – P & W had 74 houses

Legal obstacles – privacy considerations – addresses/NMIs as identifiers

Unsupervised clustering

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



# 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**<br>APS: CE3<br>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

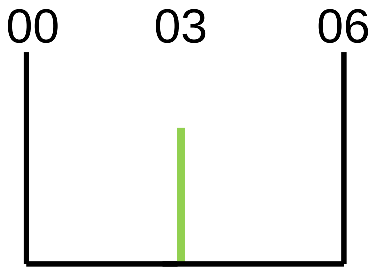
Training on ECMWF data – "**Copernicus Climate Data Store"** data (hourly, worldwide – large data)

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**”)

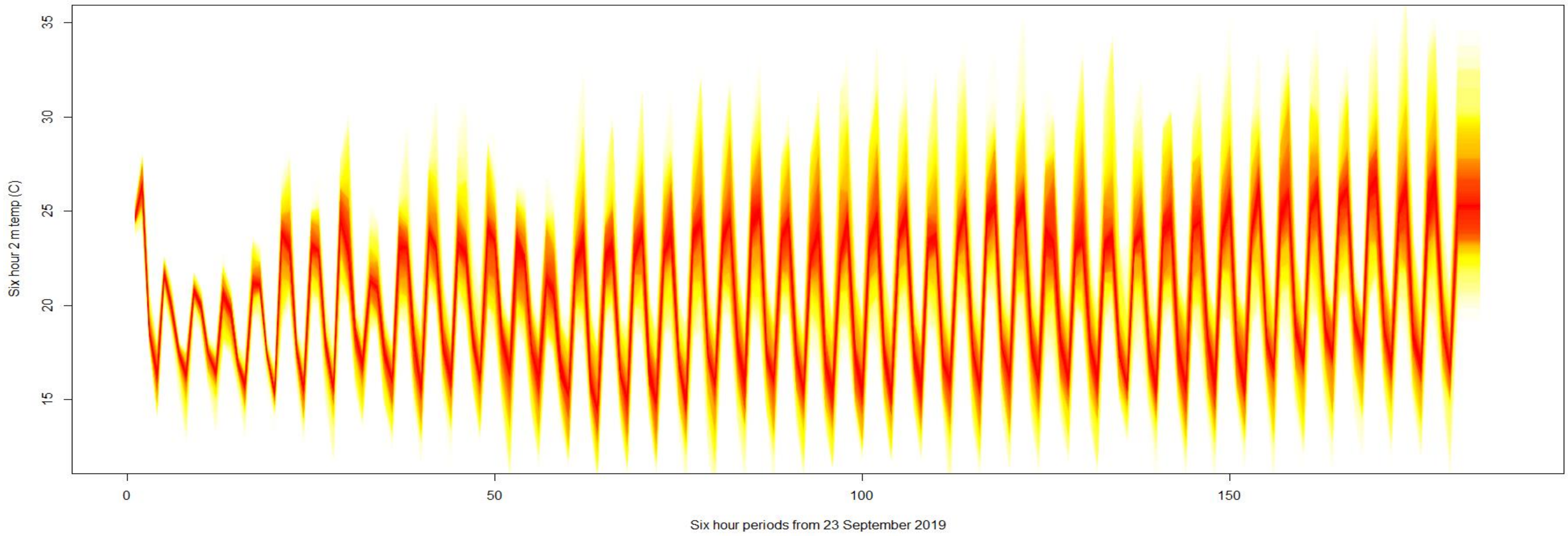
Using quantiles of output as forecast (0 to 100 per cent)

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

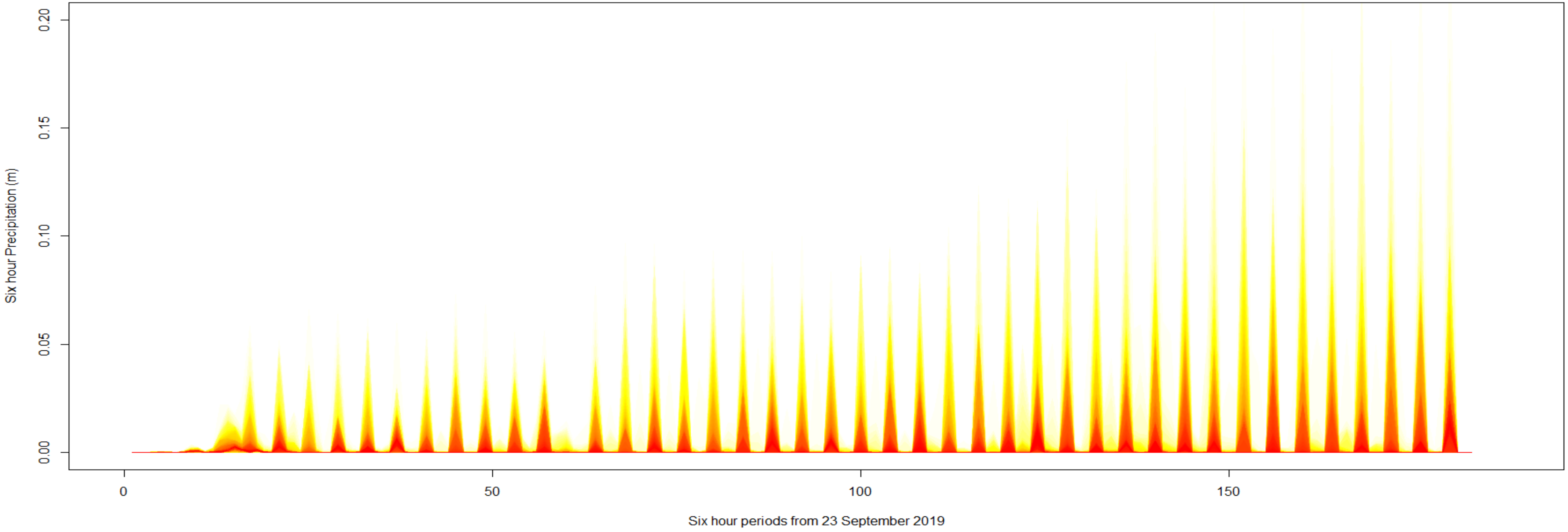
Gaps in demand data availability

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

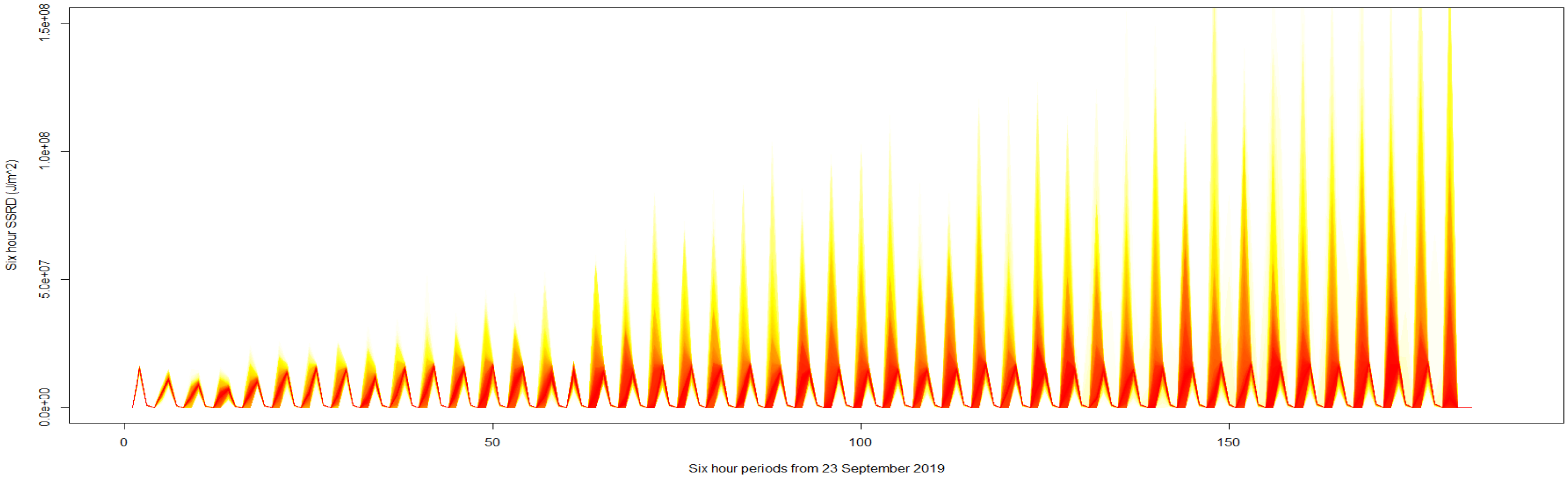
# Temperature fan plot



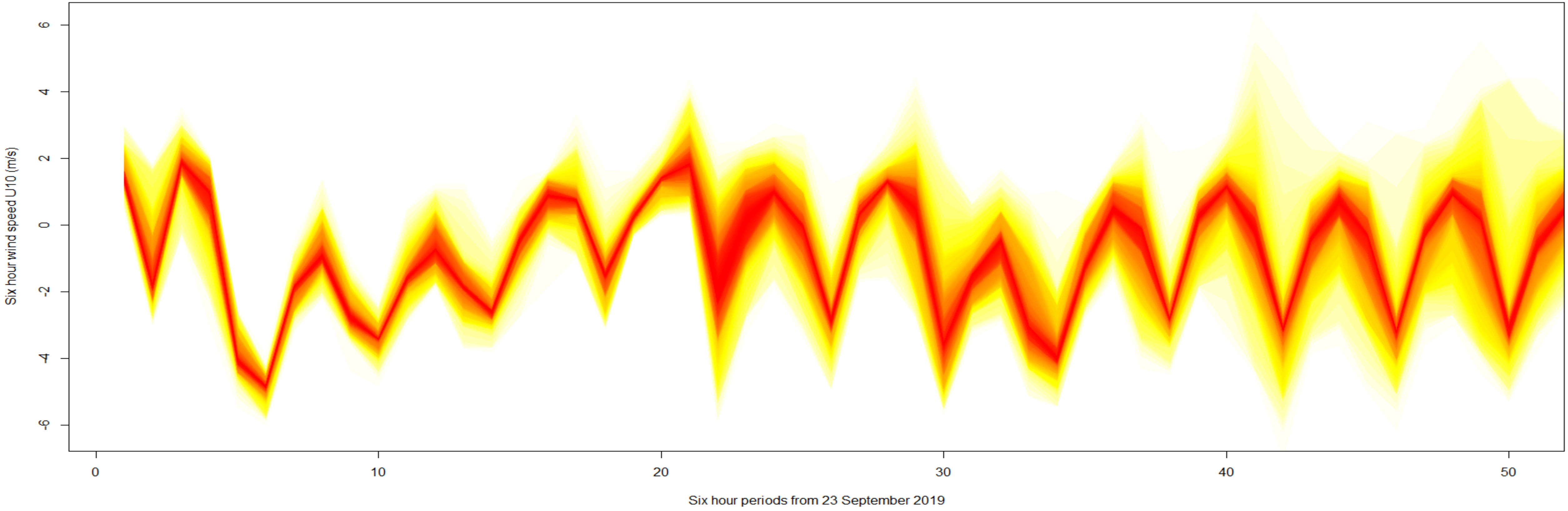
# Precipitation fan plot



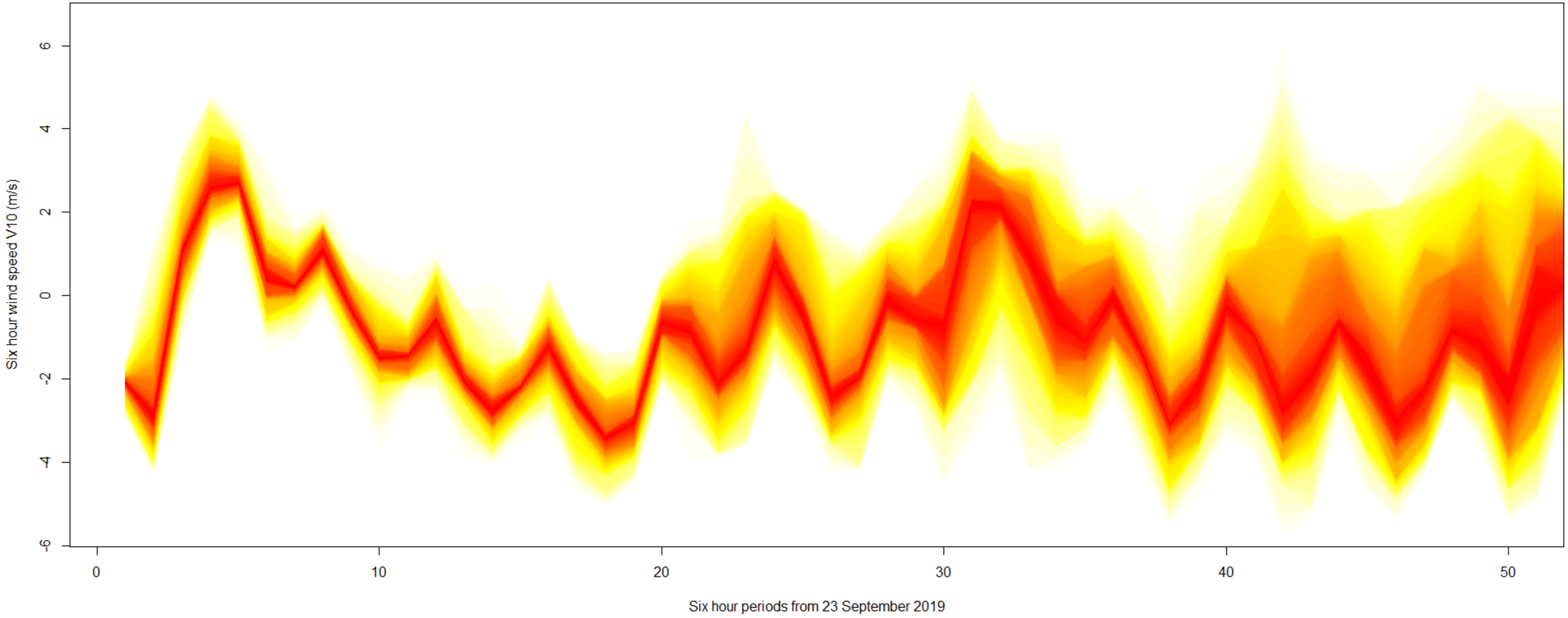
# Surface solar radiation downwards fan plot



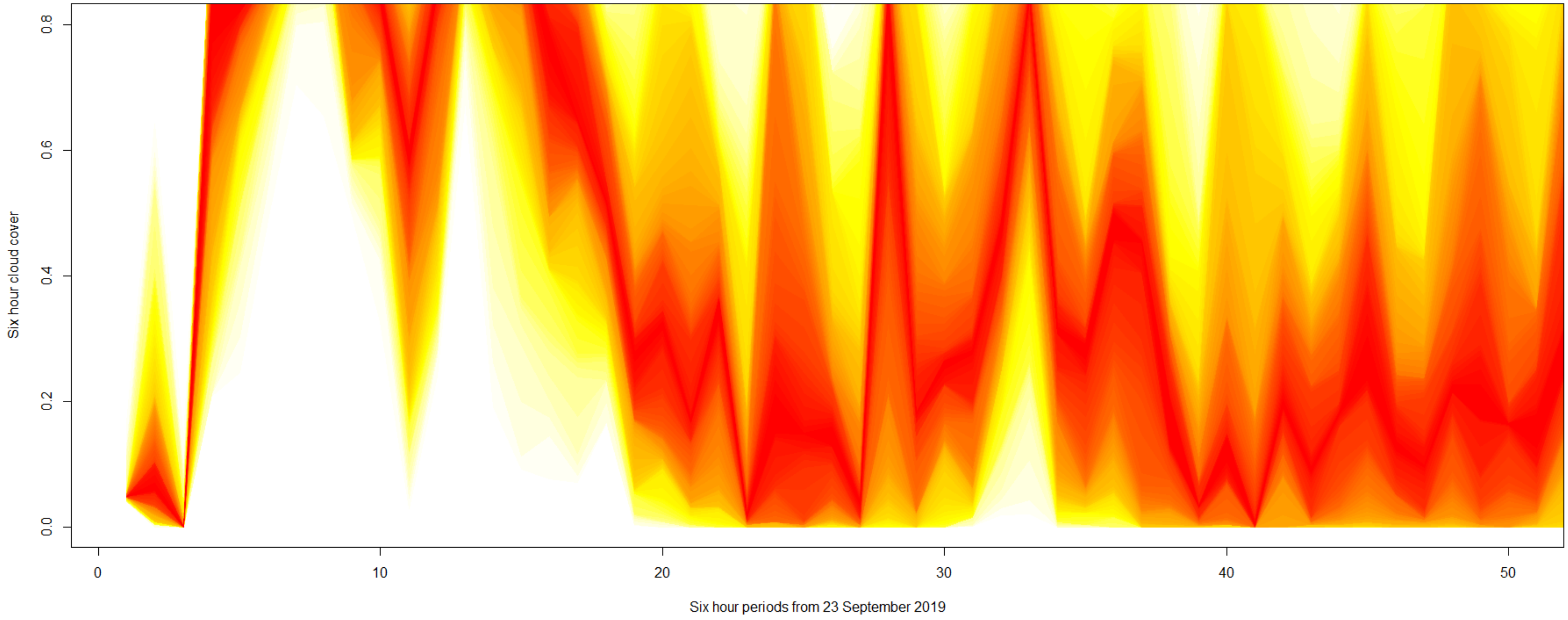
# Six hour wind speed U10 (m/s) zonal east-west



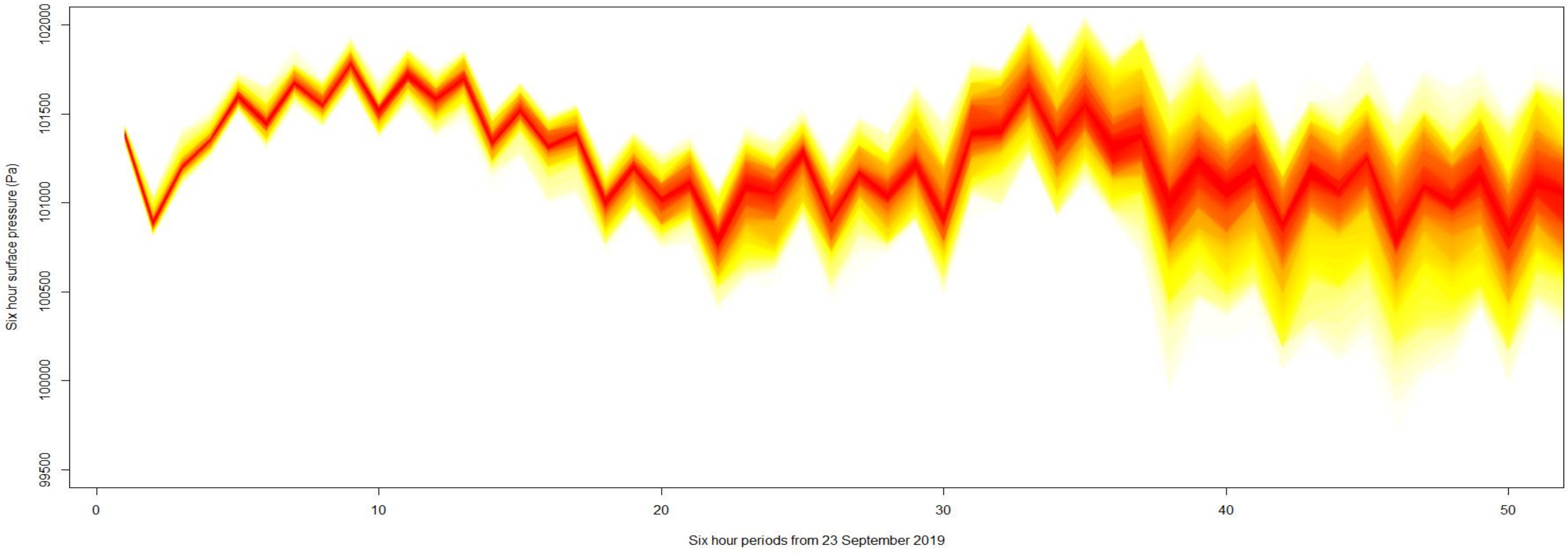
# Six hour wind speed V10 (m/s) meridional north-south



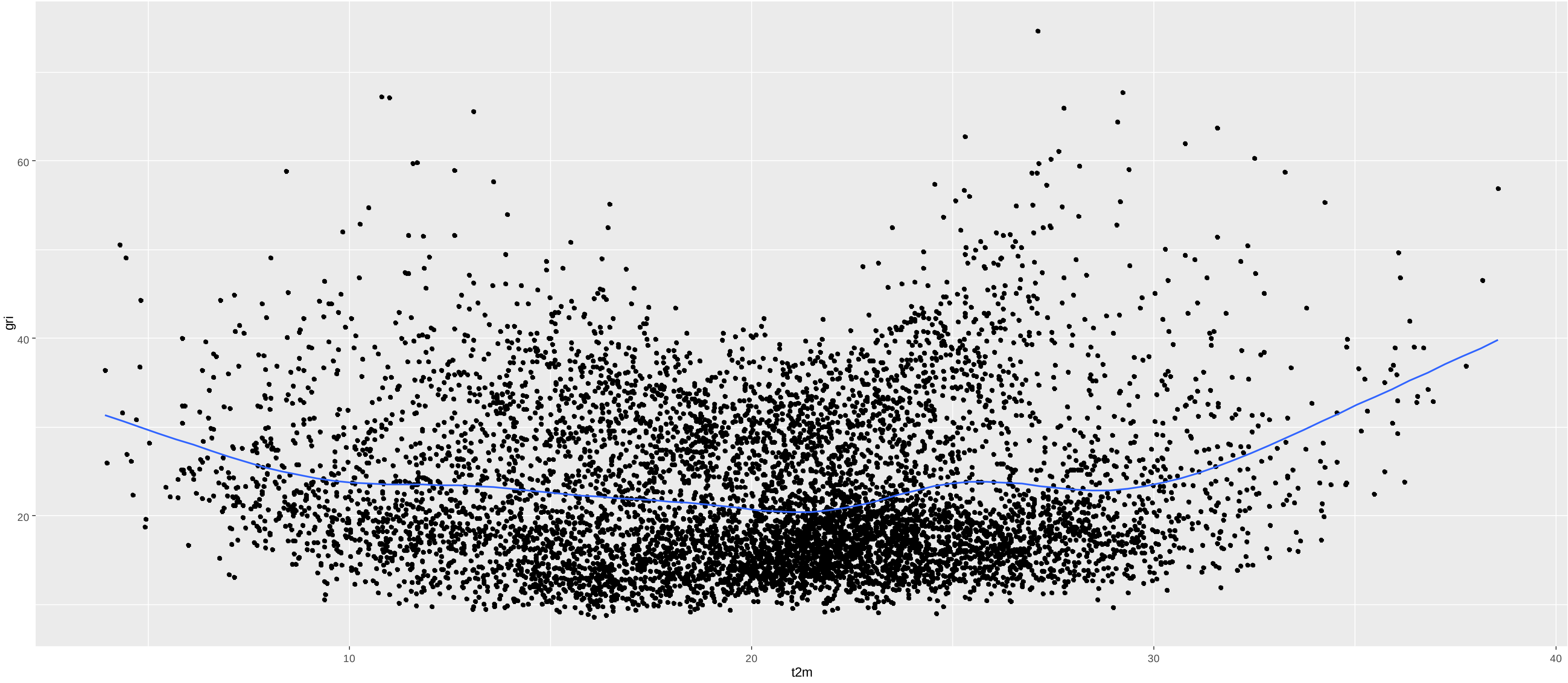
# Six hour cloud cover



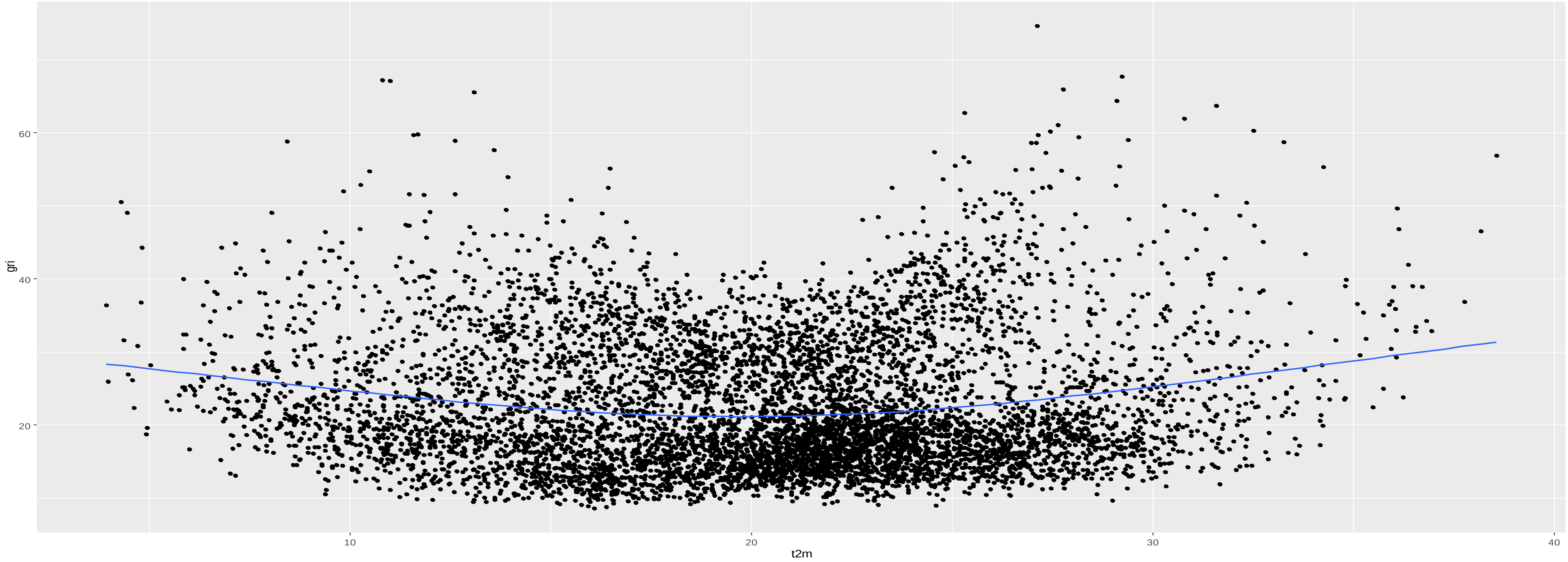
# Six hour surface pressure (Pa)



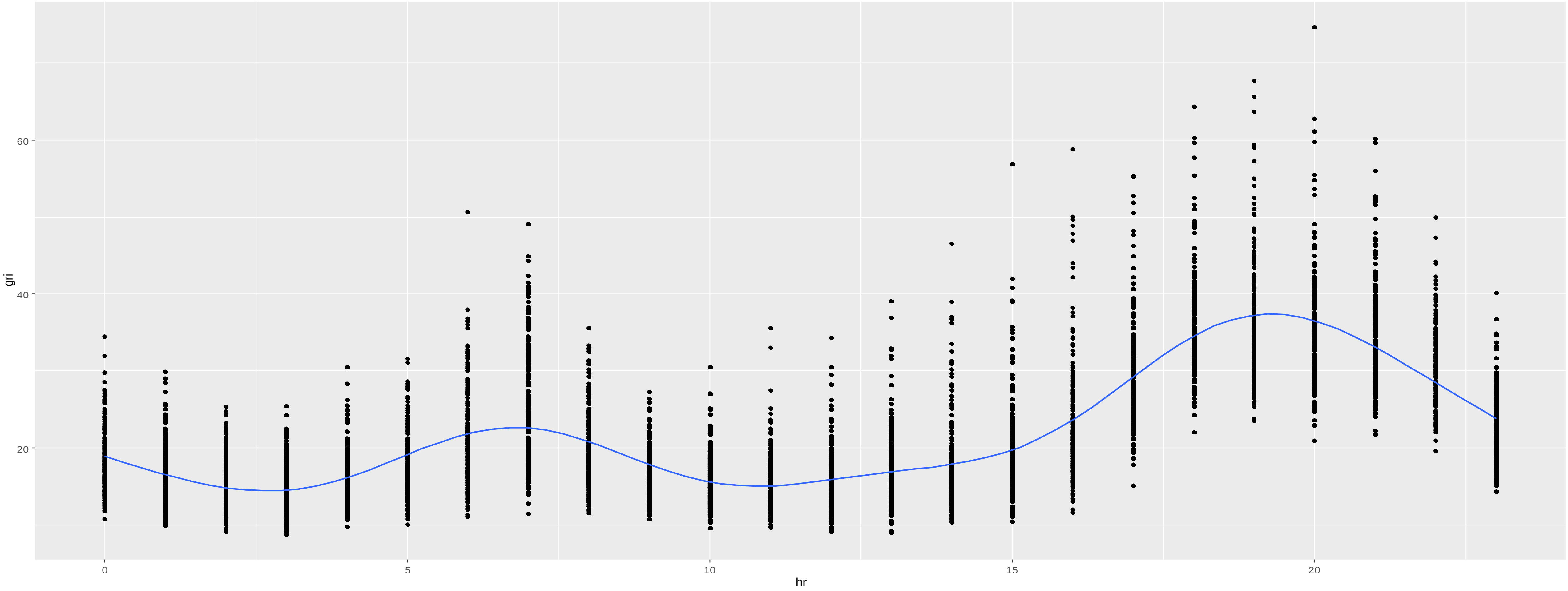
# Spline regression – Temperature vs Grid Import



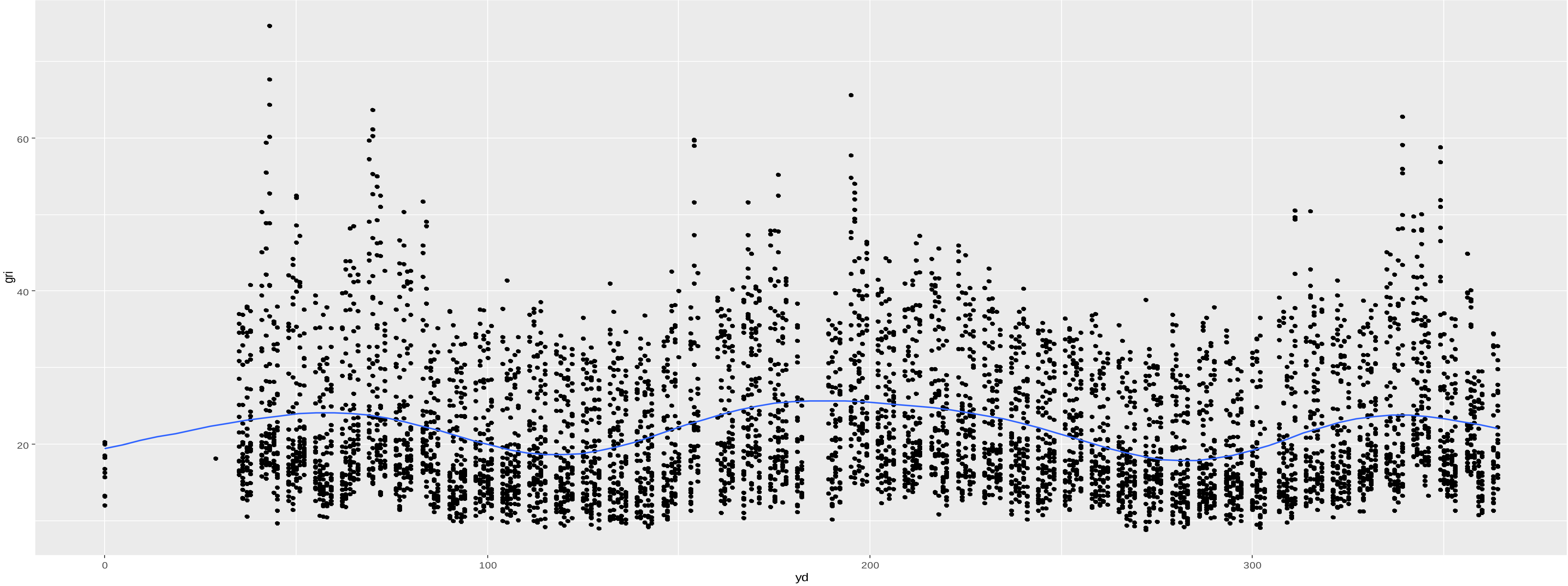
# Spline regression – Temperature vs Grid Import (3 knots)



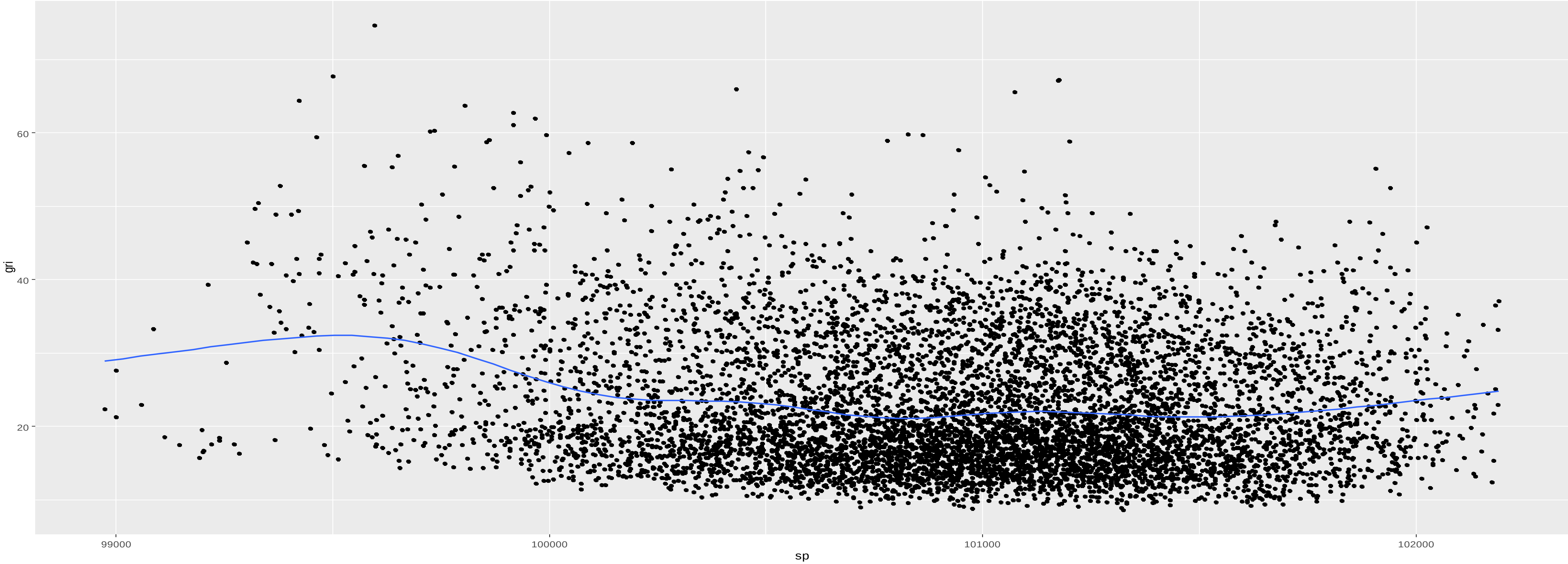
# Spline regression – Hour vs Grid Import



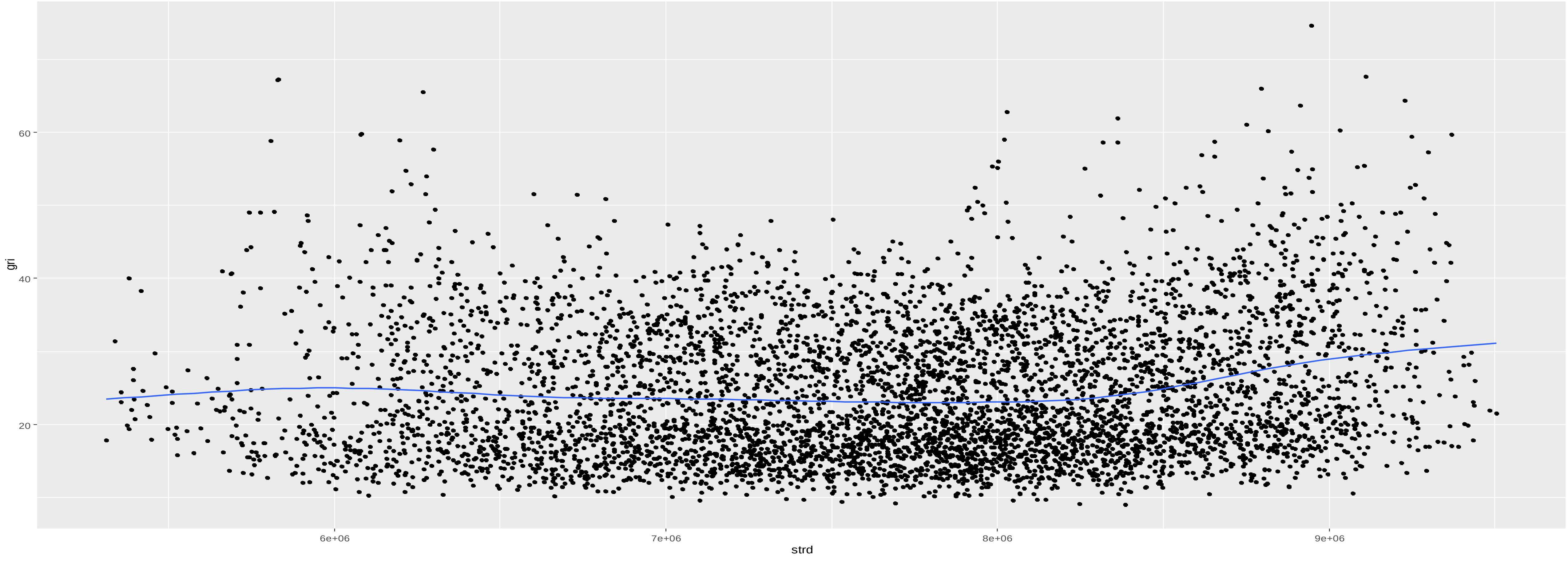
# Spline regression – Julian Date vs Grid Import



# Spline regression – Surface pressure vs Grid Import



# Spline regression – Surface Thermal Rad Down vs Grid Import



# Load and Solar Forecasting

After we work out the most useful variables (weekend/weekday model split, hour, temperature, Julian date, surface thermal radiation down, surface pressure) – for Mitchelton - work on visualizations.

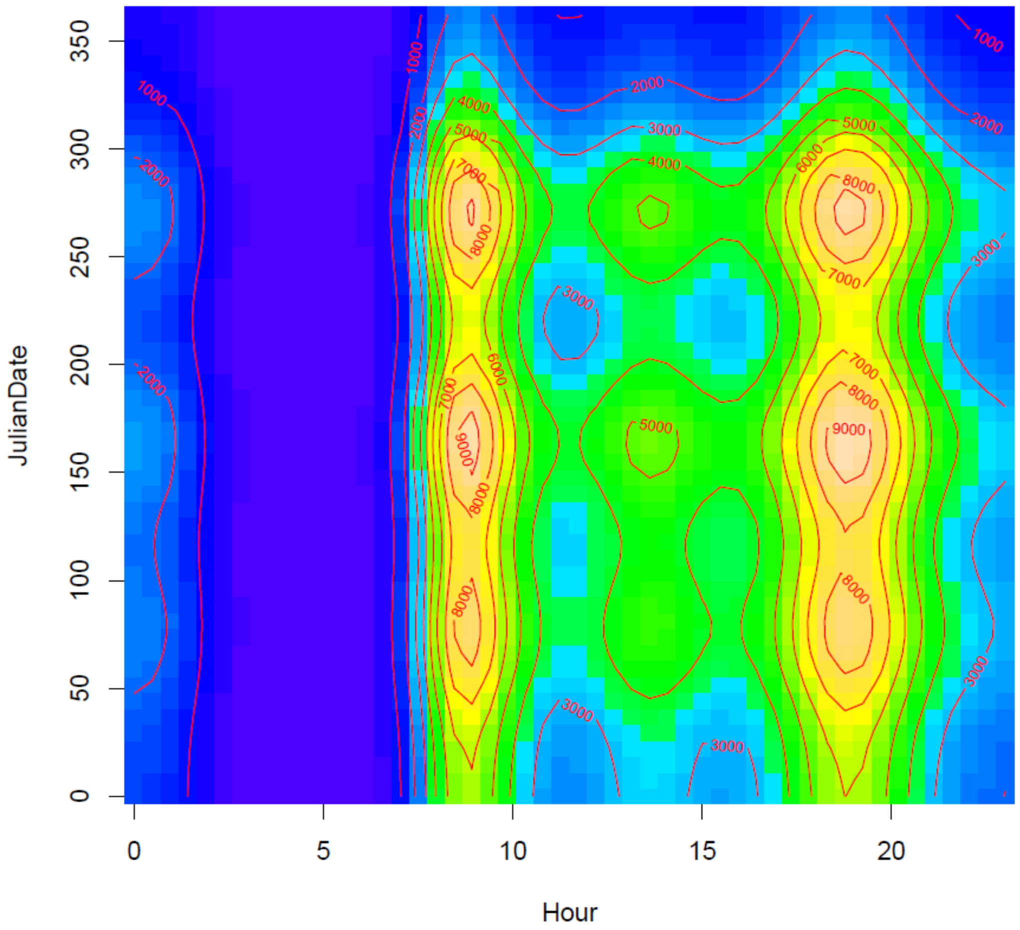
Mitchelton houses have 50% PV so solar is a very significant variable. At higher level, expect wind, precipitation etc to be significant variables. Also lagged demand, public holiday variables become significant.

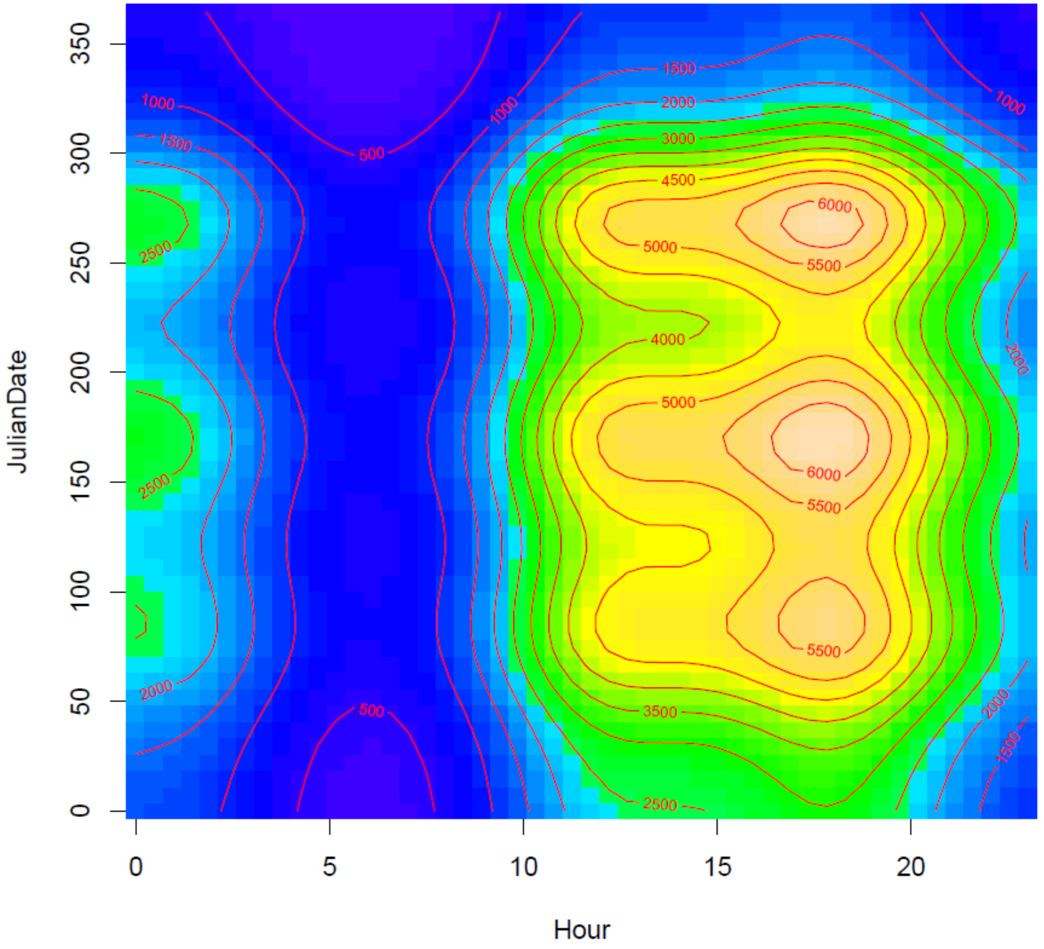
- Visualizing pairwise interactions
- Generalized additive models (mgcv package) – Simon Wood
- Visualize the fitted values with a contour plot

Start off with some bike sharing demand visualizations. Same variables – much more data.

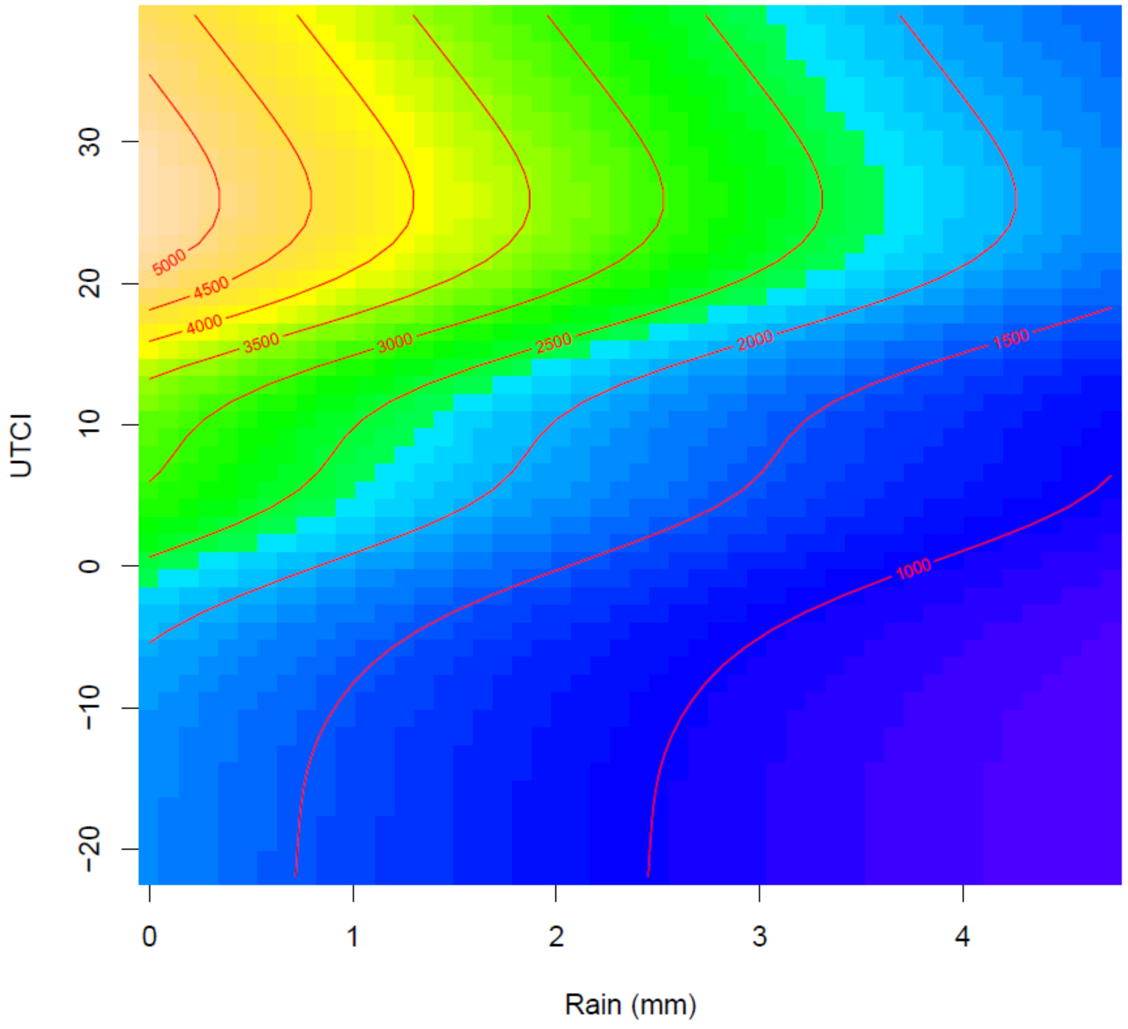
Human energy and transport behaviour is highly predictable.

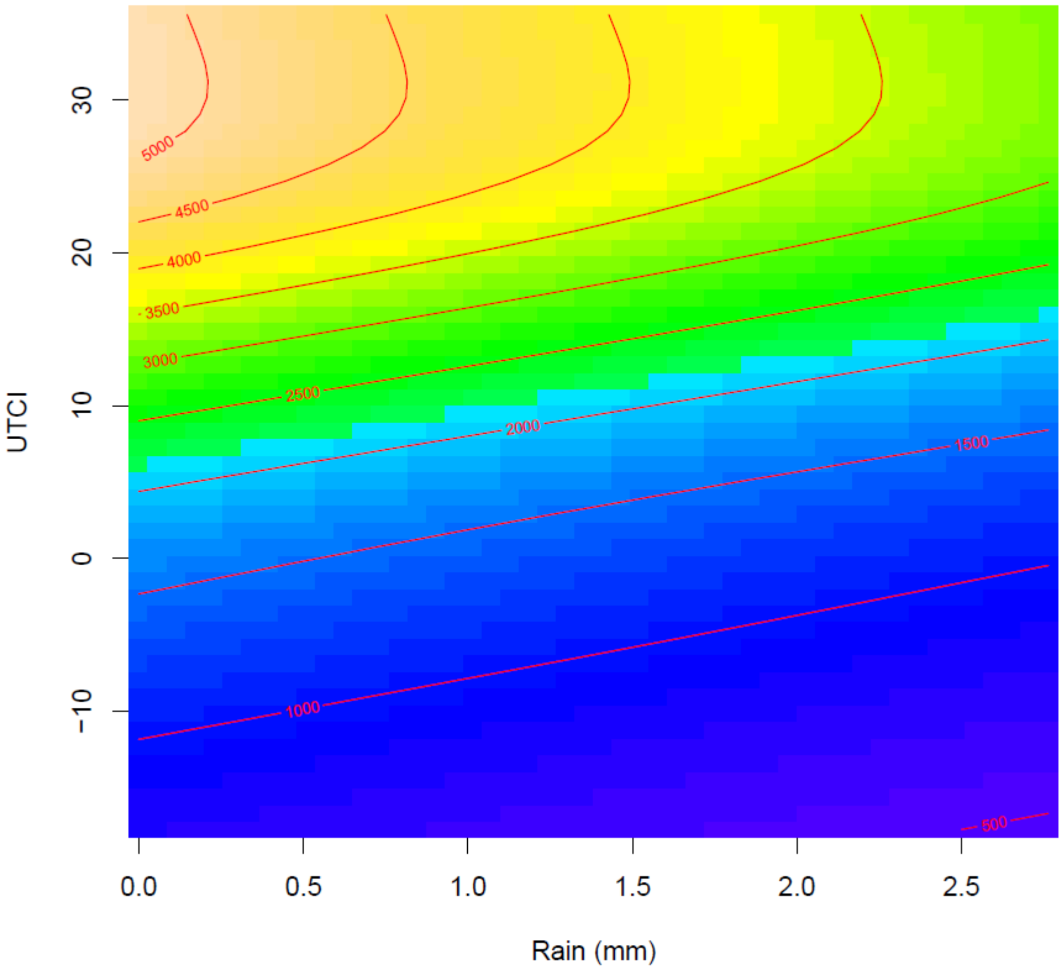
# Paris usage by hour and Julian date – weekday/weekend model



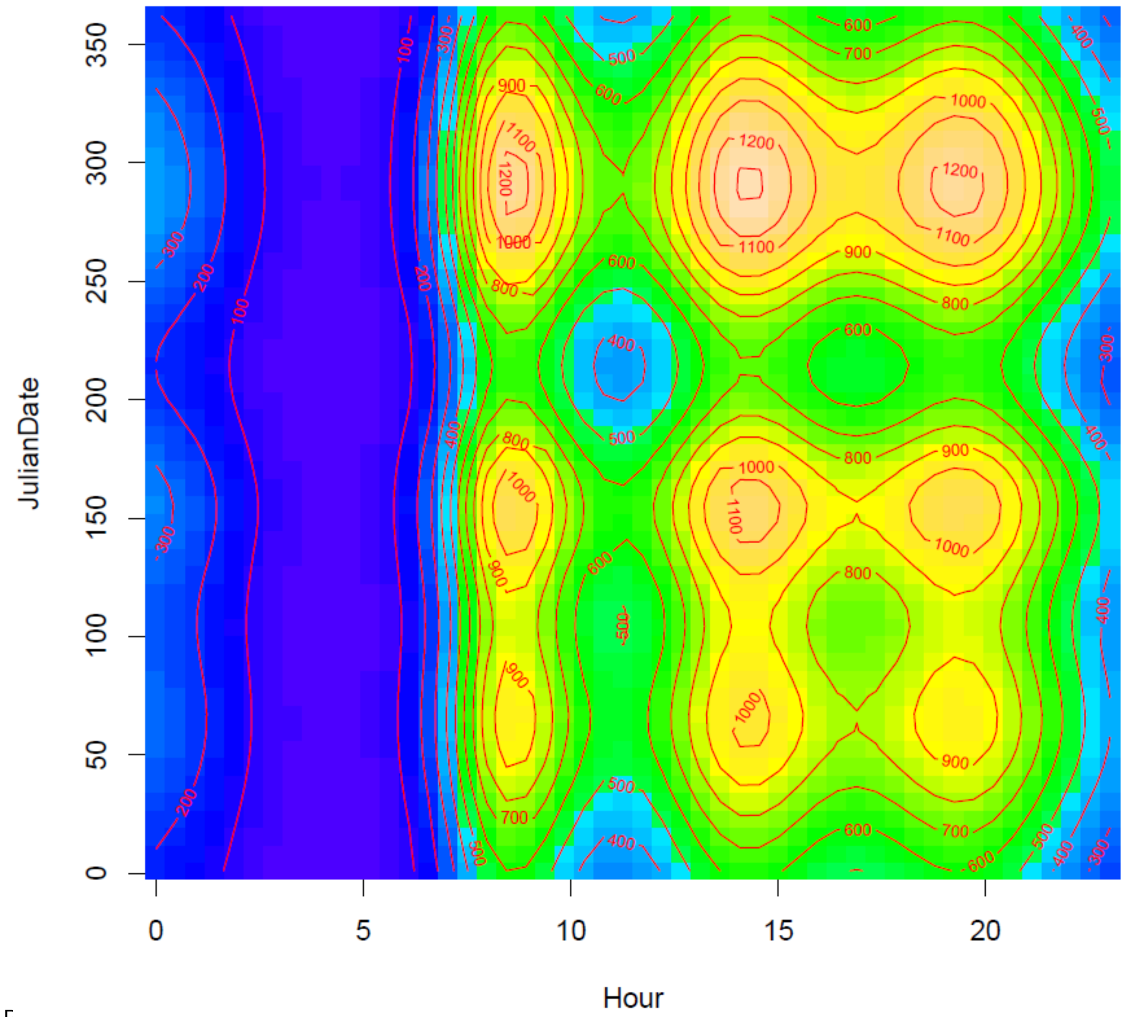



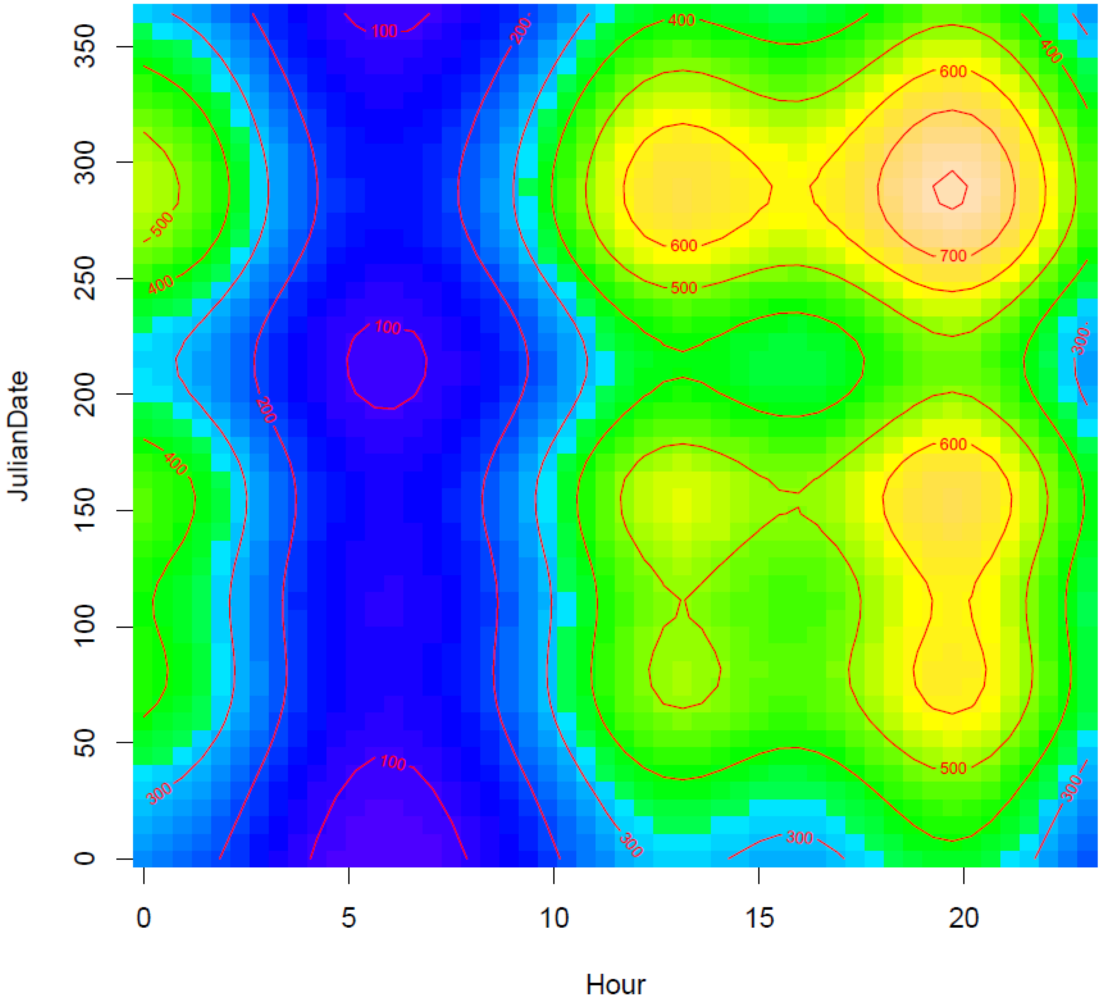
# Paris usage by rain and UTCI – weekday/weekend model



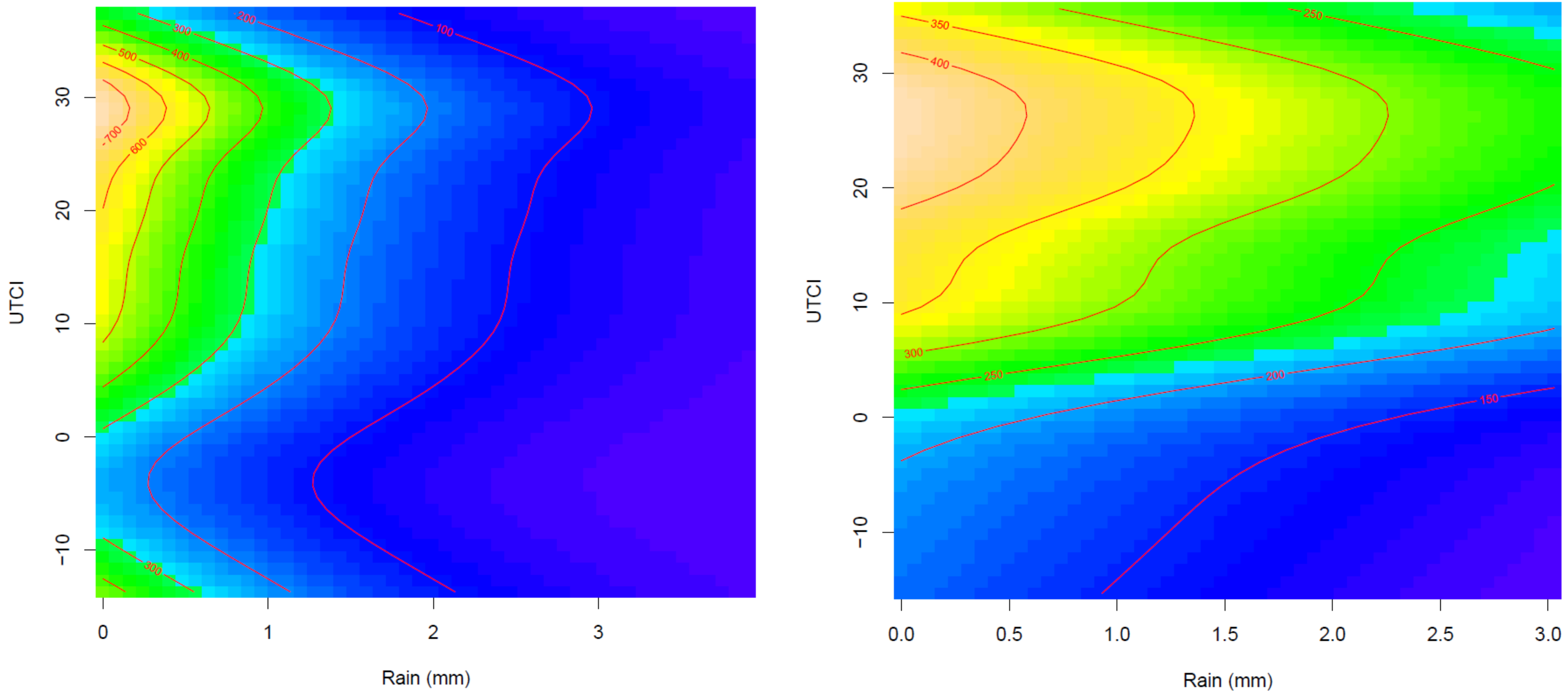



# Valencia usage by hour and Julian date – weekday/weekend model

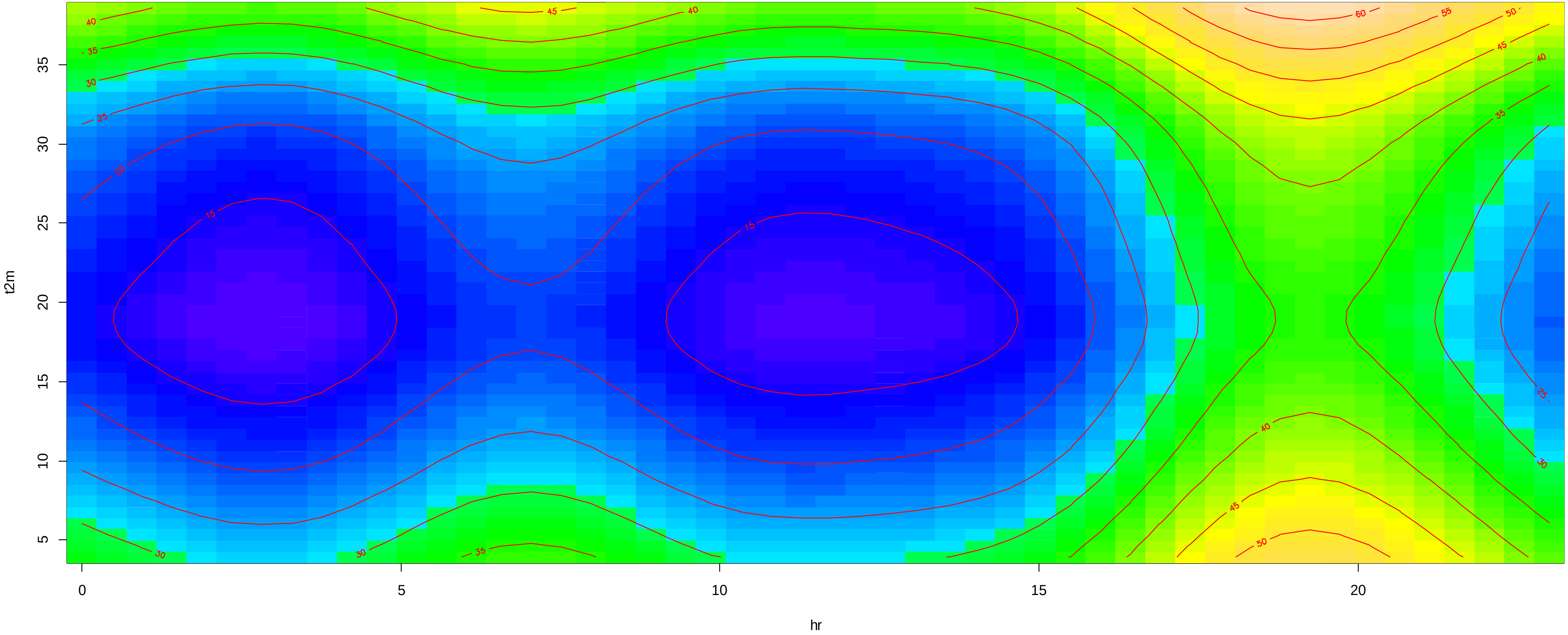


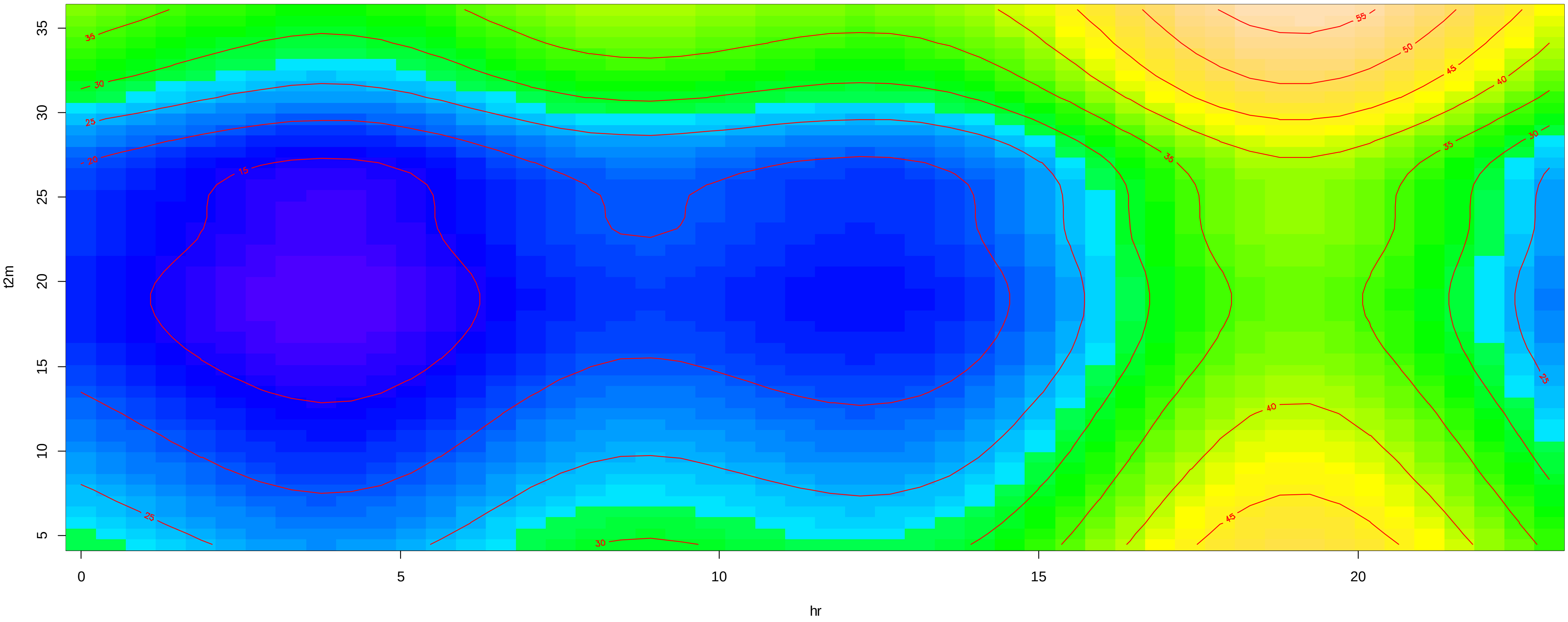
# Valencia usage by rain and UTCI – weekday/weekend model



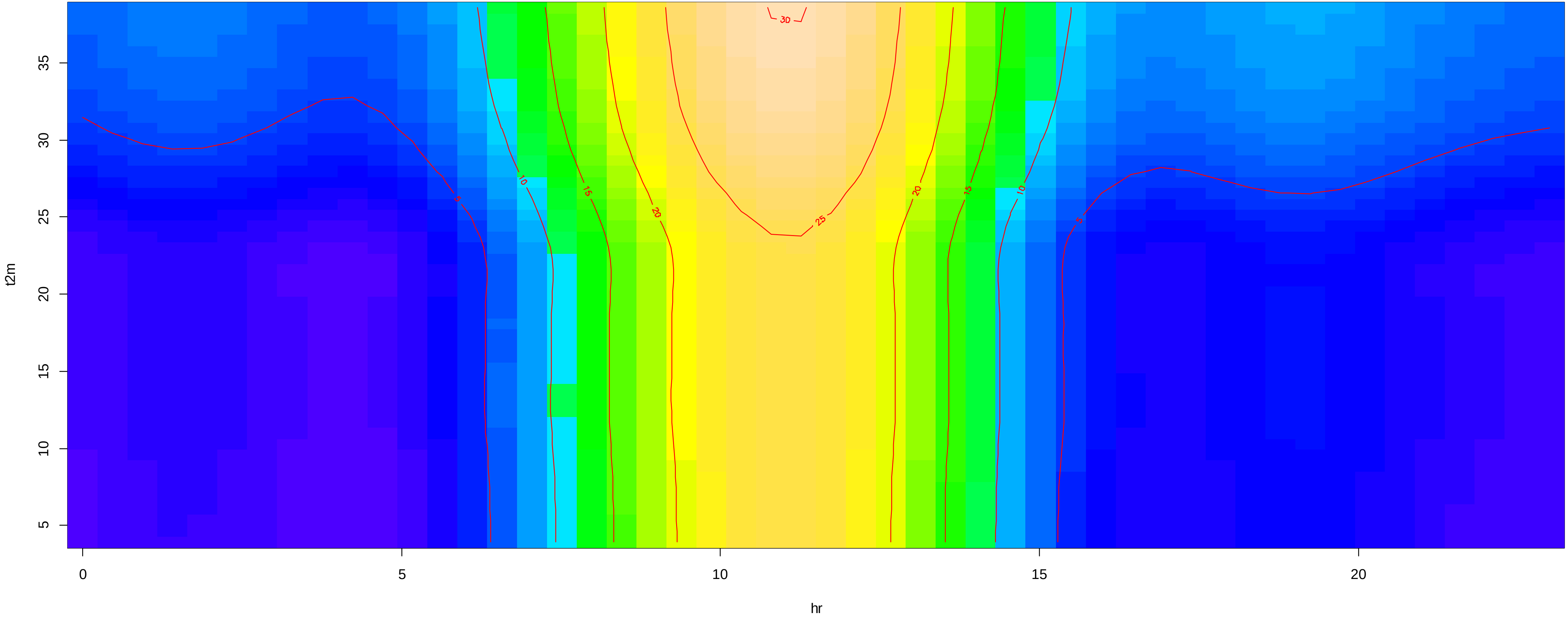
# Weekday: Grid Import by Hour and Temperature



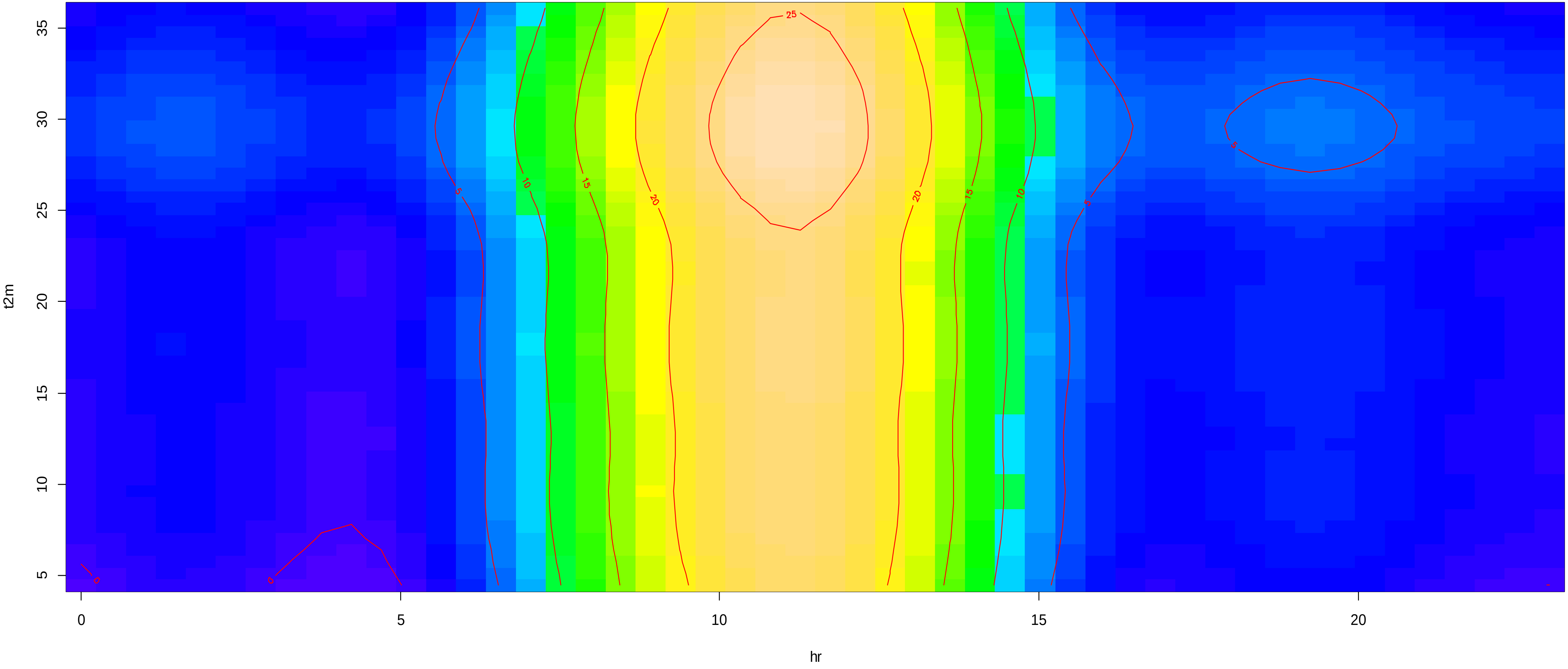
# Weekend: Grid Import by Hour and Temperature



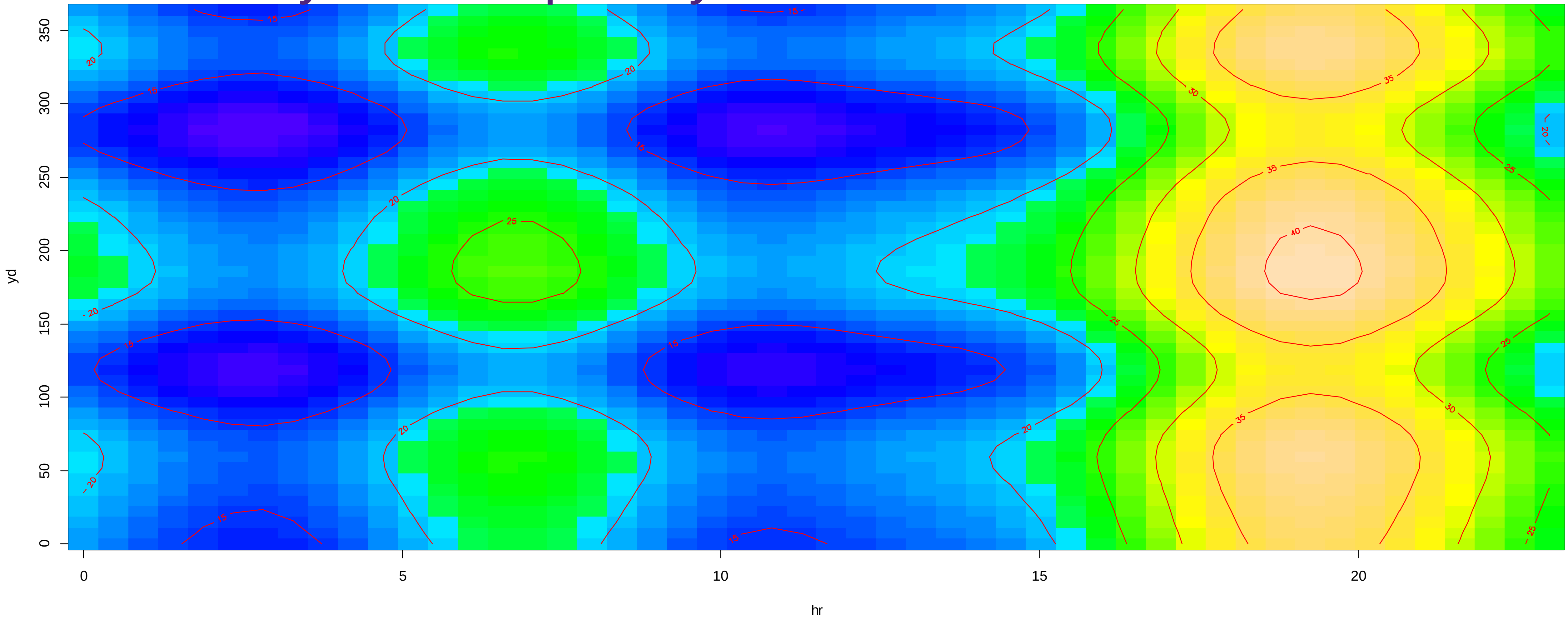
# Weekday: Grid Export by Hour and Temperature



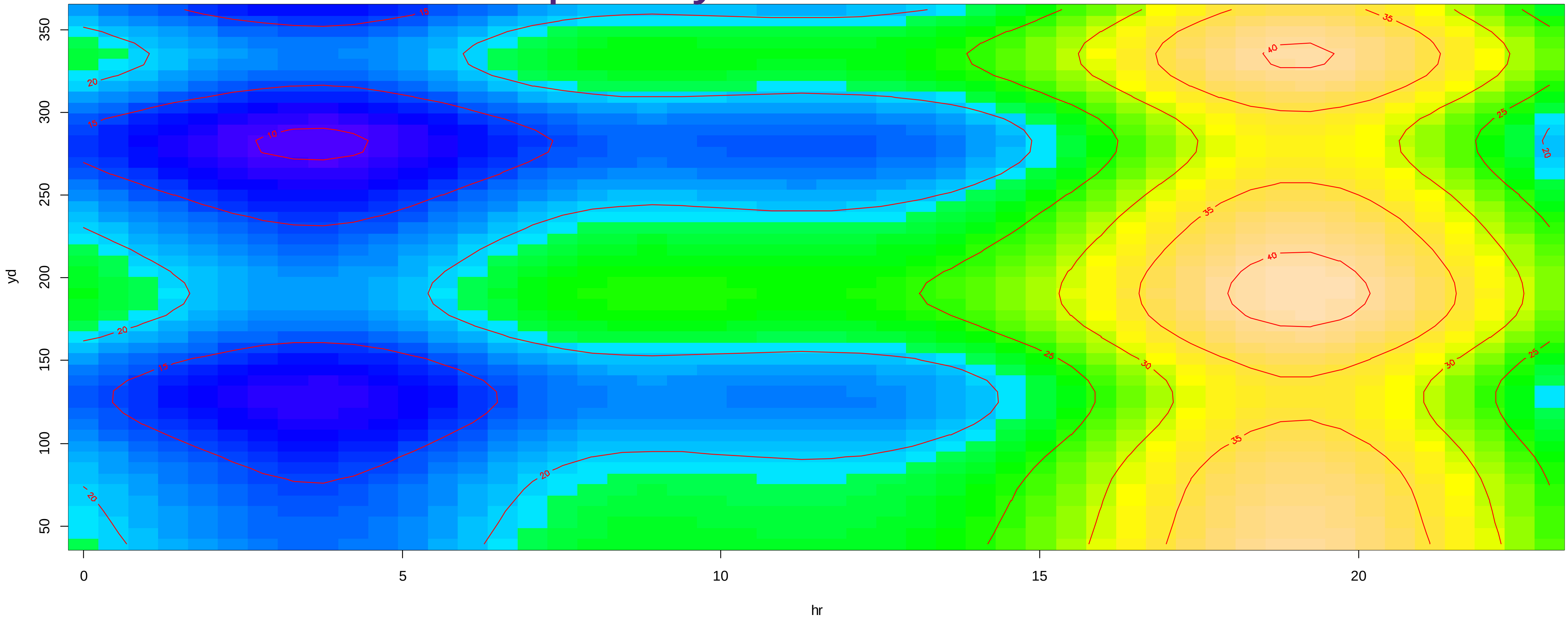
# Weekend: Grid Export by Hour and Temperature



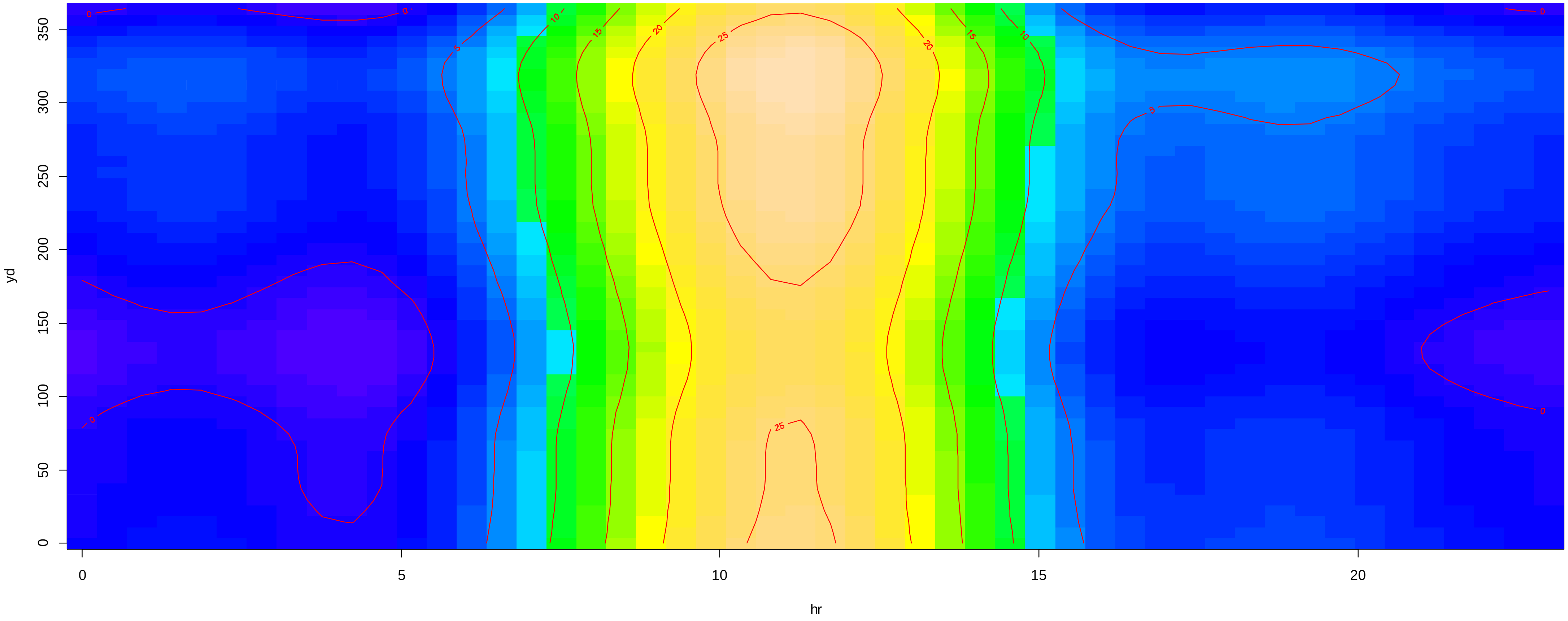
# Weekday: Grid Import by Hour and Julian date



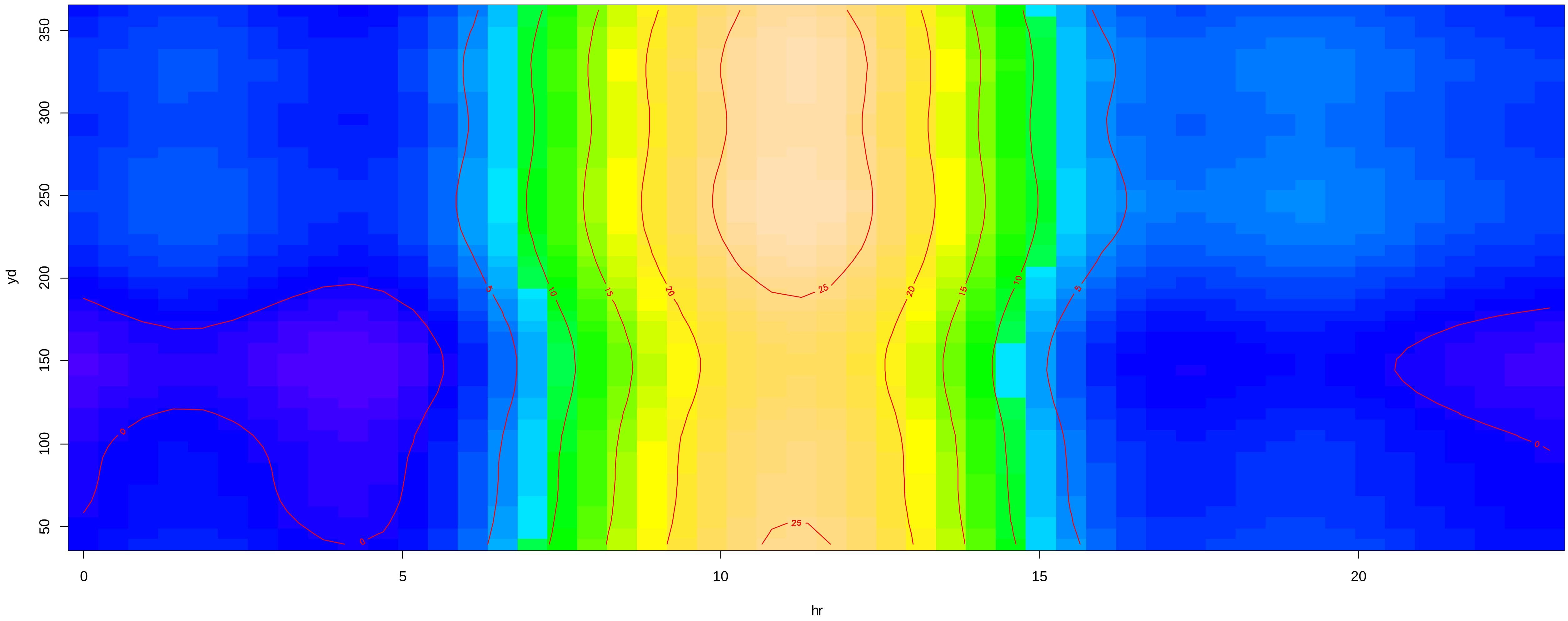
# Weekend: Grid Import by Hour and Julian date



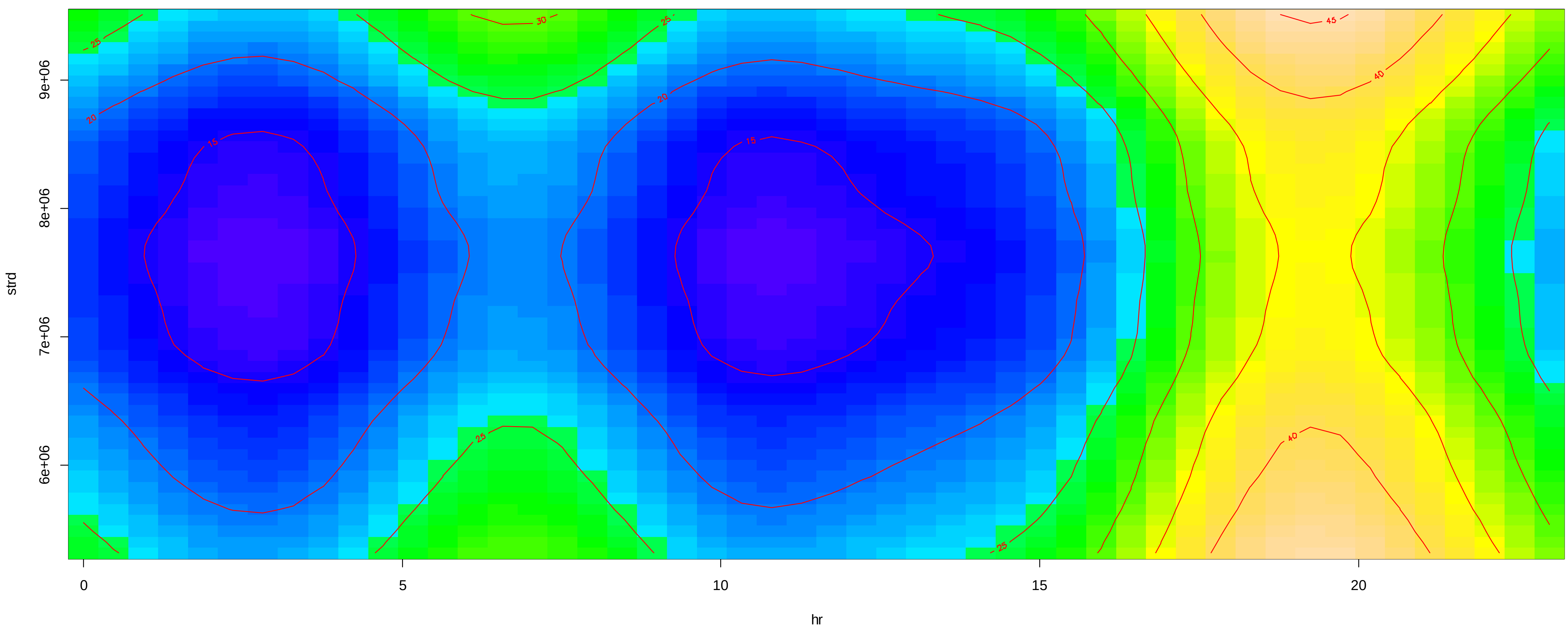
# Weekday: Grid Export by Hour and Julian date



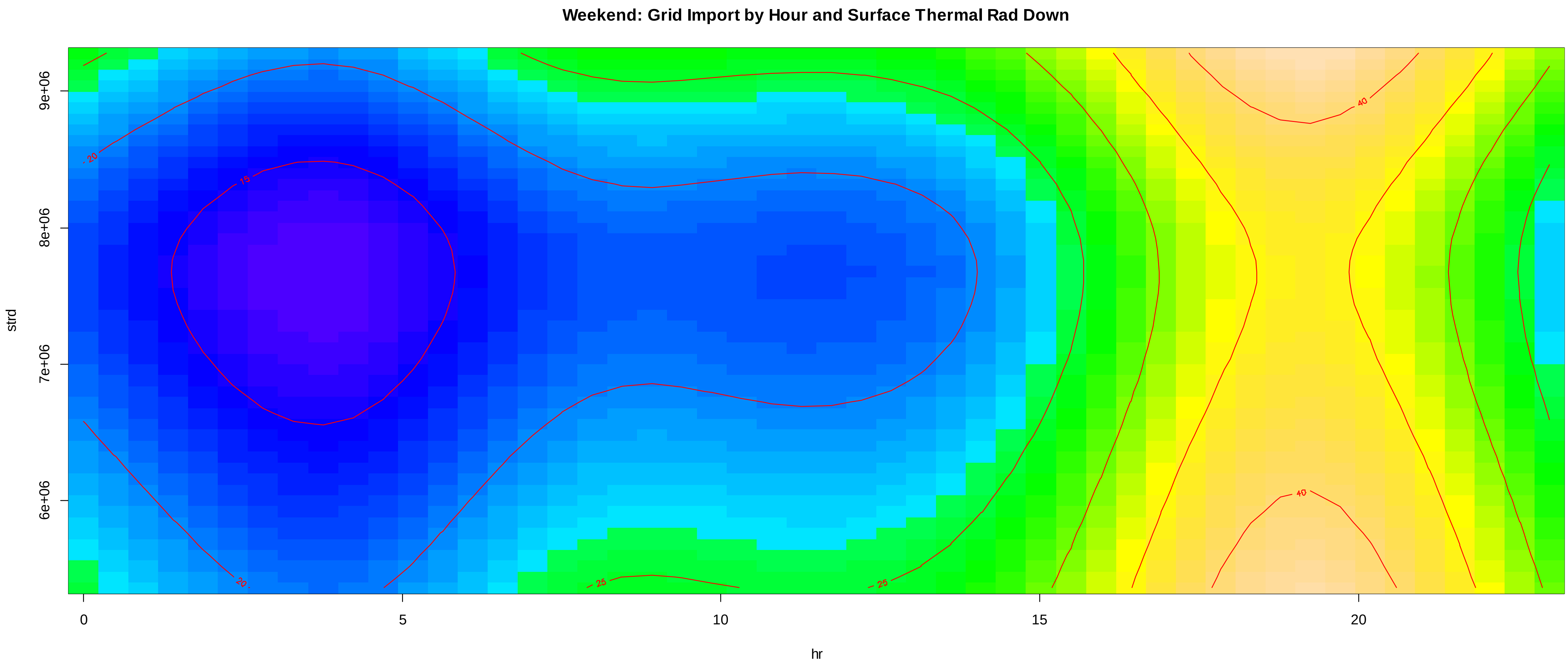
# Weekend: Grid Export by Hour and Julian date



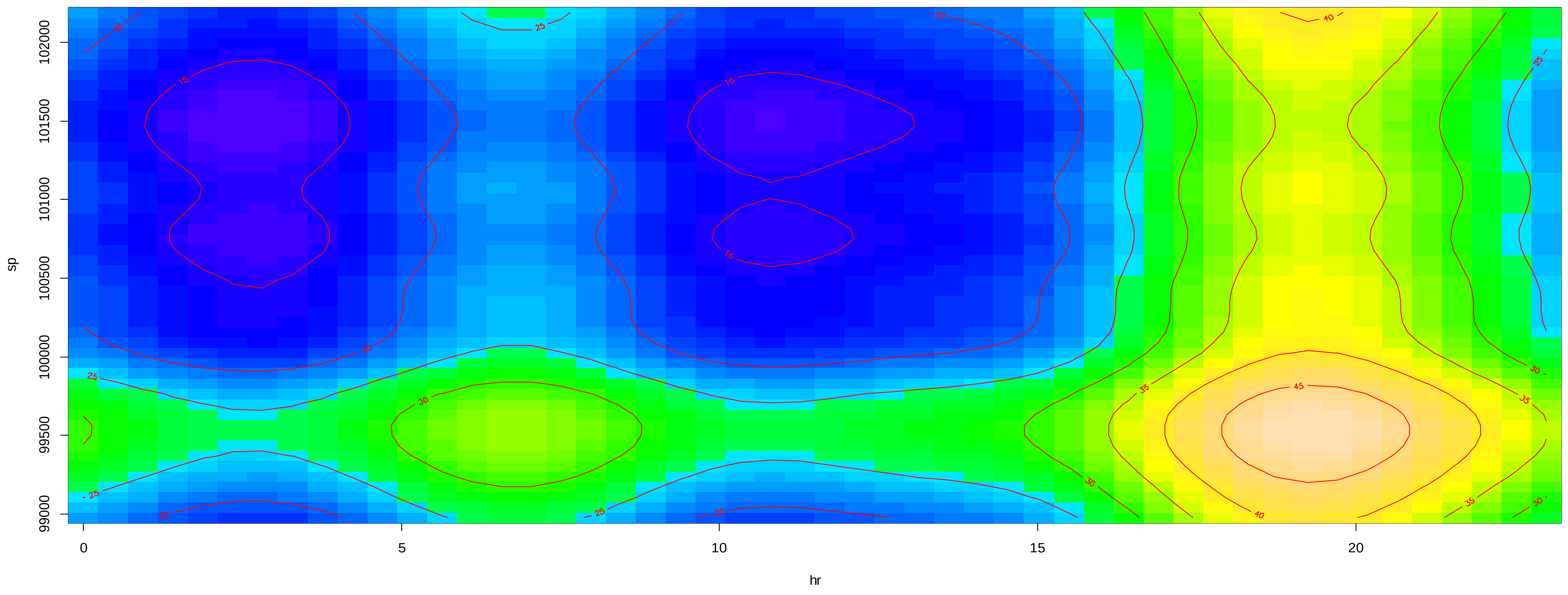
# Weekday: Grid Import by Hour and Surface Thermal Rad Down



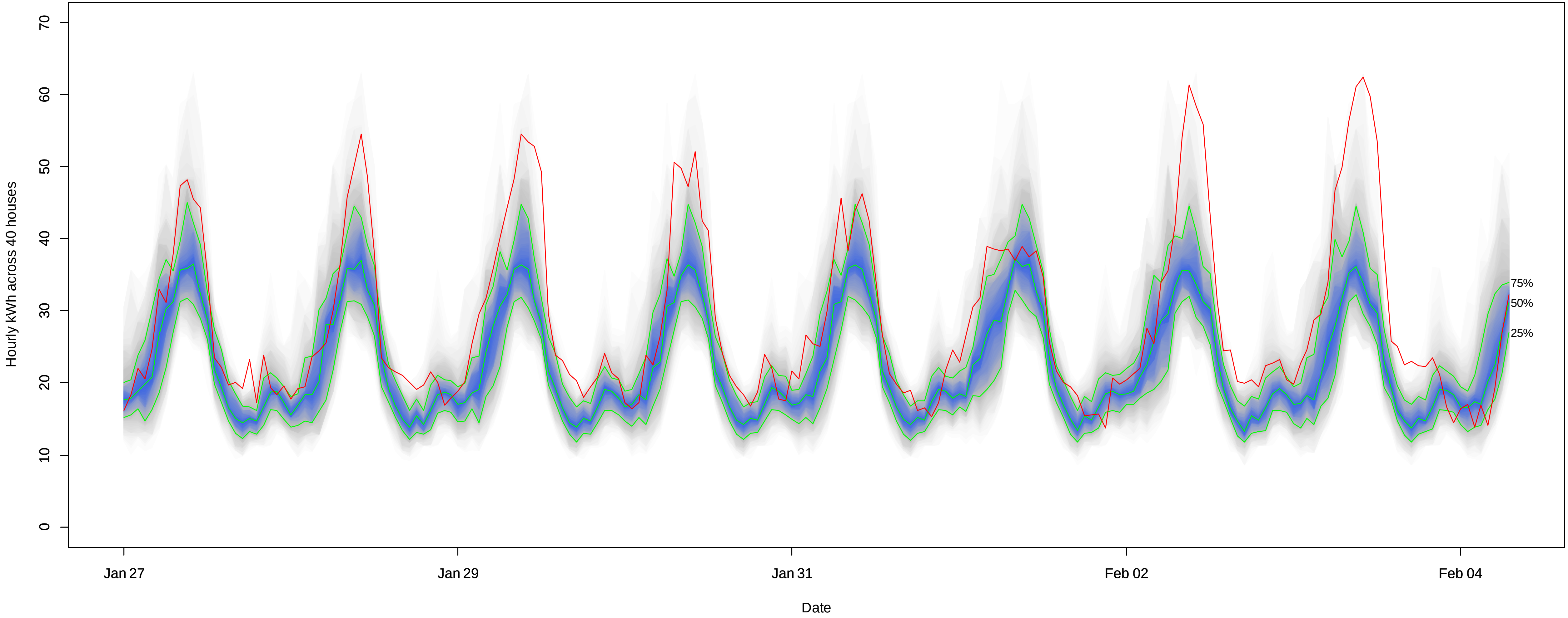
# Weekend: Grid Import by Hour and Surface Thermal Rad Down



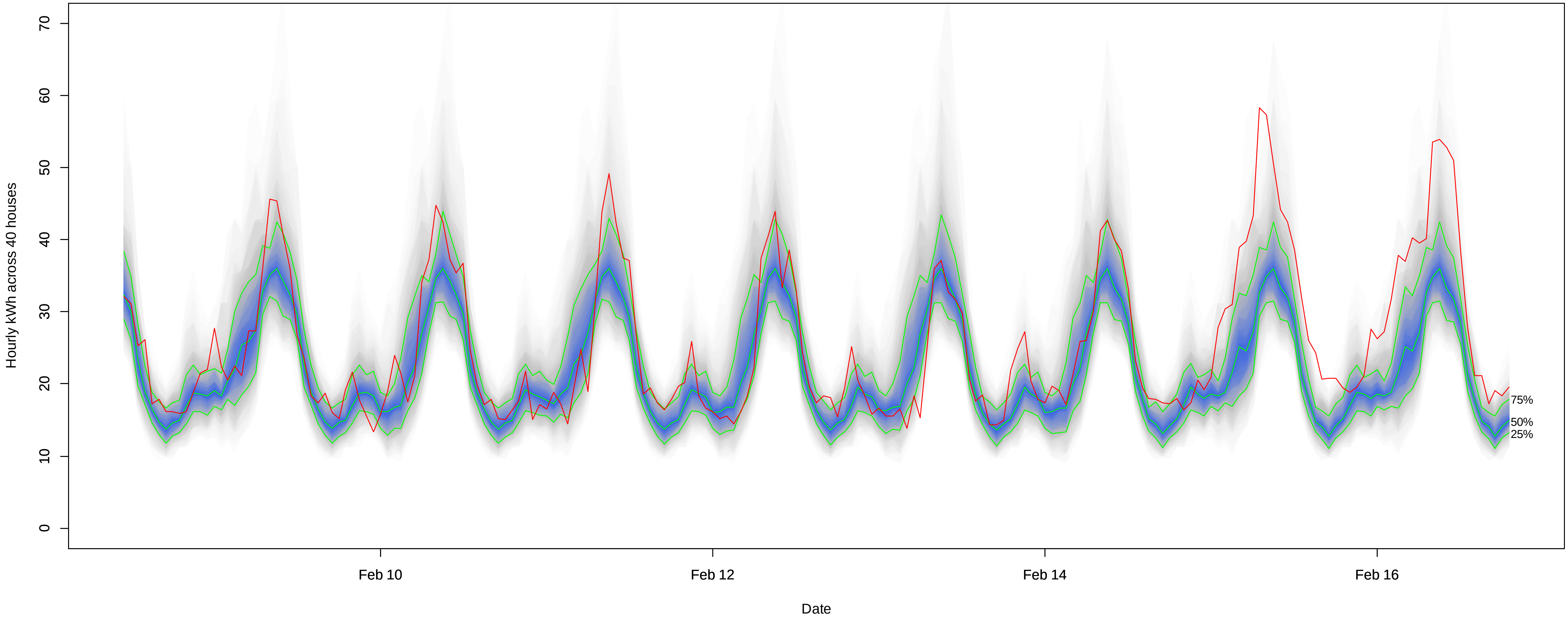
# Weekday: Grid Import by Hour and Surface Pressure



# Forecast from 27 Jan – quantile RF, 40 houses



# Forecast from 10 Feb – quantile RF, 40 houses



# 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
- State monitoring of networks
- Visualizations – e.g. different kinds of customer, projected to two dimensions
- Better understanding of metering and how it relates to billing (parallel with Metering and Settlement)
- Forecasting system interruptions due to severe weather conditions

THE UNIVERSITY OF QUEENSLAND AUSTRALIA | CREATE CHANGE

# Thank you

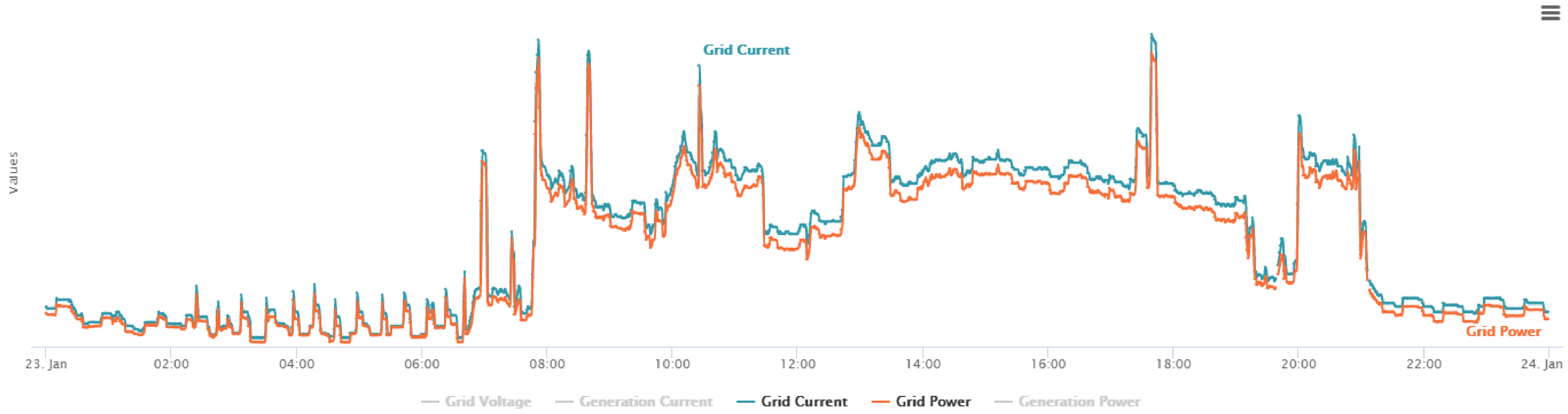
Richard Bean, Research Fellow
Centre for Energy Data Innovation
r.bean1@uq.edu.au

facebook.com/uniofqld

Instagram.com/uniofqld

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

# NILM - Spot the appliances



# Exploring Smart HWS

**Data acquisition system workflow:**

individually addressable, stores data login

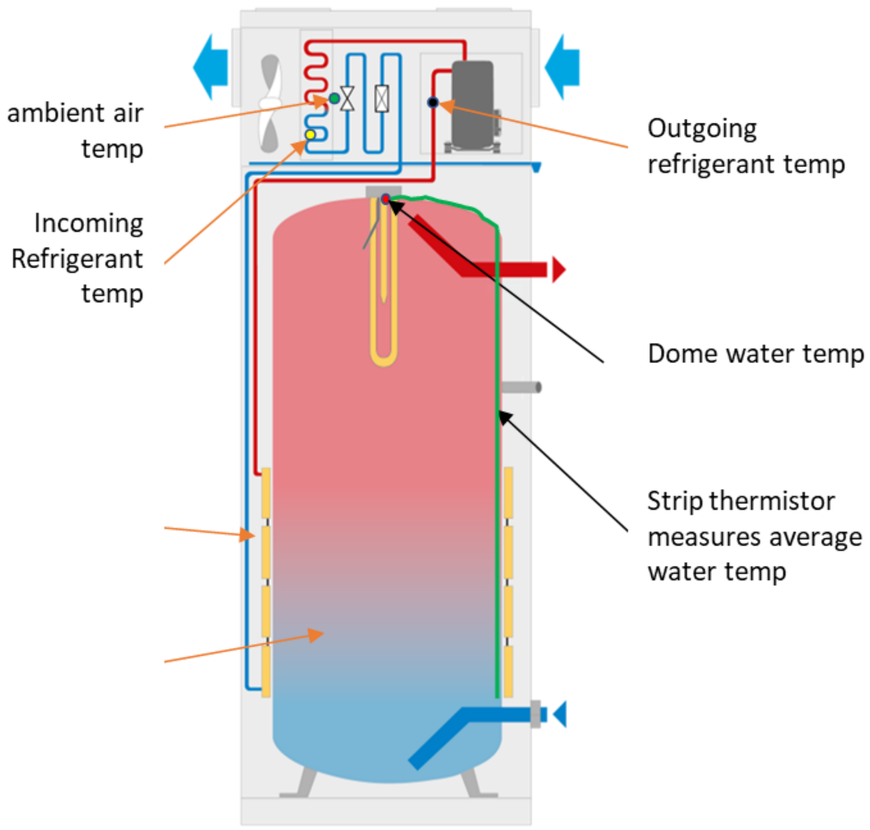
**Data reporting rate:**

**1** minute (adjustable)

**Data acquisition entities:**

| | | | |
|---|---|---|---|
| 1* | 1st setpoint temp | 10 | Error menu |
| 2* | 2nd setpoint temp | 11 | Hot gas switch temp |
| 3* | Loading/heating level | 12 | Air inlet temp T1 |
| 4 | Cylinder size | 13 | Hot gas temp T1 |
| 5 | Mixed water volume | 14 | Evaporator temp T1 |
| 6 | Electric element | 15 | Stored temp diff |
| 7 | Compressor | 16 | Dome temp |
| 8 | Fan | 17 | Integral temp |
| 9 | Defrosting | | |

**1,2,3 entities are configurable*




- 1 installed at a family of 4 (in use)
- Test lab under construction that can mimic (different household behaviours and different environmental conditions)

# Hot Water Usage Behavior Description

1. Understand customer usage patterns
2. Understand how HWS responds to those usage patterns
3. Identify opportunities for individually addressable
    - solar sponge
    - Demand Management