Forecasting and Optimizing a Microgrid for the IEEE-CIS Technical Challenge

Dr Richard Bean

University of Queensland

Ingkarni Wardli Room B.18

AUPEC 2022, University of Adelaide

3.10pm, Monday 26 September

Personal Background

- Redback Technologies (2016-2019)
  - Smart solar inverter manufacturing
  - Forecasting house solar generation, optimising cost
- University of Queensland (2019-now)
  - Centre for Energy Data Innovation https://cedi.uqcloud.net/
- Project SHIELD
  - Collection of 1-minute energy data from ~20,000 homes in Qld
  - Aimed at improving dynamic operating envelope for solar in distribution networks

Competition

- IEEE Computational Intelligence Society (CIS)
- Forecasting and Optimization Technical Challenge 2021
  - Running from July to November 2021
- 6 buildings at Monash with 6 solar systems; 2 batteries

Competition Requirements

- Develop an
  - optimal battery schedule
  - an optimal lecture schedule – *recurring activities and once-off activities*

based on predictions of future values of energy demand and production.

- With input data:
  - **Energy consumption data** (15 min) to September 2020
  - **Solar production data** (15 min) to September 2020
  - **Daily weather data** from Australian Bureau of Meteorology (daily solar, max and min temperatures, rain) – three sites
  - **Hourly weather data** from European Centre for Medium Range Weather Forecasting (ECMWF)
  - **Electricity price data** from AEMO (30 minute)

ABSTRACT

- From a machine learning point of view, the provided data poses an interesting time series prediction problem, with
  - *multiple seasonality,*
  - *use of external data sources (weather, electricity price)*
  - *the opportunity for cross-learning across time series on two different prediction problems (energy demand and solar production).*
- Then, from an optimization point of view, uncertainty in the inputs needs to be addressed together with a couple of constraints, to achieve a good solution.

Requirements and Assessment

Phase 1

- Optimally schedule a battery and timetabled activities (lectures) for the month of October 2020.
- In real life, the battery scheduling would usually happen on a daily basis, with day-ahead forecasting.
- Phase 1 - public leaderboard where participants submit forecasts and the leaderboard shows the evaluation of the forecasts (MASE error rate and cost)

Phase 2

- Data for October 2020 is released to the participants, and they are now asked to perform the same forecasting and optimisation exercise for November 2020.
- Now, only minimal feedback is provided to the participants about the quality of their submissions. Solely Phase 2 of the competition is relevant to determine competition winners and prizes.
- The 3 main competition prizes will be awarded to the schedules that lead to the **lowest cost on the Phase 2 test set. ($US7000, $US5000, $US3000)**
- An additional prize will be awarded to the team that achieves the most accurate forecasts on Phase 2. **($US2000)**

OBJECTIVE FUNCTION

Objective

For a feasible schedule, we compute the objective value in terms of the cost of the schedule, which is to be minimized. The cost of the schedule depends on three parts:

- The total energy cost computed against the wholesale price e_t,
- The peak load tariff taken over the whole month,
- The value of the once-off activities scheduled $d_i \in \{0, 1\}$, whether in or out of office $o_i \in \{0, 1\}$.

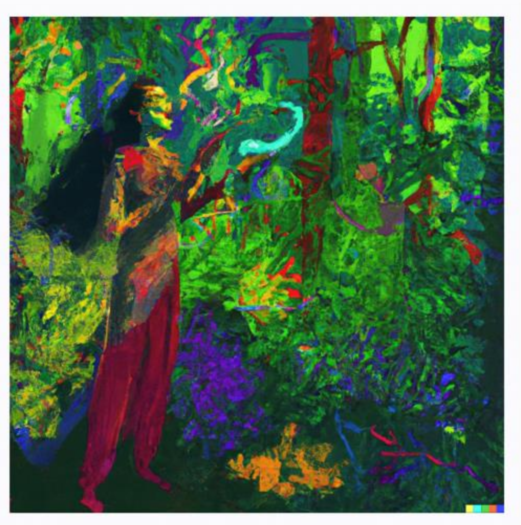
The objective O is computed as follows:

$$O = \sum_t \frac{0.25 l_t e_t}{1000} + 0.005(\max_t l_t)^2 - \sum_{a_i} (d_i \cdot (\text{value}_i - o_i \text{penalty}_i))$$

Becomes an MIQP (Mixed Integer Quadratic Program)

What worked in the past?

- Global Energy Forecasting Competition
  - GEFCOM – three editions 2012, 2014 and 2017
    - 2012 – hierarchical load forecasting and wind power
    - 2014 – hierarchical load, wind energy, price, and solar
    - 2017 – hierarchical probabilistic load
- Used to inform Redback model (2016-2019)
- Random forests; quantile regression forests where quantiles required



ELSEVIER

International Journal of Forecasting
Volume 30, Issue 2, April–June 2014, Pages 357-363

Global Energy Forecasting Competition 2012

Tao Hong [a], Pierre Pinson [b], Shu Fan [c]

ELSEVIER

International Journal of Forecasting
Volume 32, Issue 3, July–September 2016, Pages 896-913

Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond

Tao Hong [a], Pierre Pinson [b], Shu Fan [c], Hamidreza Zareipour [d], Alberto Troccoli [e], Rob J. Hyndman [c]

ELSEVIER

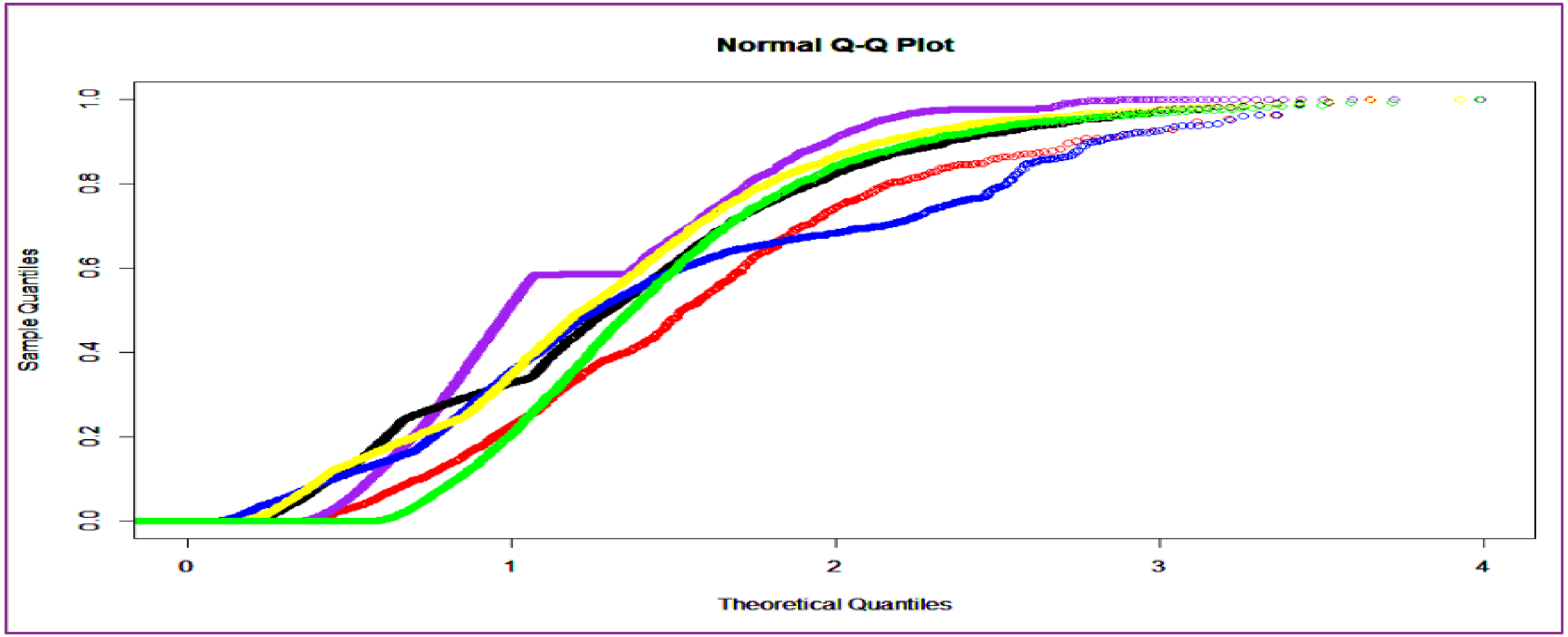
International Journal of Forecasting
Volume 35, Issue 4, October–December 2019, Pages 1389-1399

Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting

Tao Hong [a], Jingrui Xie [b], Jonathan Black [c]

PHASE 1 FORECASTING

- Random forest libraries have the useful “variable importance” ranking
- Lagging and leading data (up to 3 hours)
- Day of week, time of day, Julian date (Fourier values)
- “Days before” temperature effect
- Data cleaning – missing data



KEY STEPS

- Quantile regression forest – forecast **median** to minimise MAE

  (i.e. sum of deviations from actual value)
- Most important parameter to tune – *"mtry"*
- Training against individual phase 1 time series (without overfitting)
- Each hour gets 4 random forests (each quarter hour)
- Choosing building start months of 2020 (Building 0,1,3,6)
- Removing building outliers
- Choosing solar start months (Solar1 has some cumulative data)
- Predictor variables: ECMWF vars lead/lag 3h, day of week, day of year etc
- Public holiday – 23 October Grand Final holiday excluded from training
- Building4 and Building5 set to median values of Oct 2020 (1 and 19 kW)

- Forecast groups of buildings and solar together with normalization (critical, but mentioned by organizers *"cross-learning across time series"*)
- Using BOM daily and ECMWF 1 hour data together (critical … is this surprising?)
- Solar0 and Solar5 thresholding hugely improves MASE (critical)

The University Of Queensland AUSTRALIA

Performance in Phase 2

- Best forecast of 9 of 12 time series (3/6 Buildings, 6/6 Solar)
- Combining all entries only improves to 0.6276
- Using AEMO 5-minute Vic demand, PVOutput allows more improvement

	Bean	Abolghasemi	SZU	EVERGi	Stratigakos	FRESNO	Best of any	1 min solar	PVOutput 2	PVOutput 7
Average MASE	**0.6460**	**0.7441**	**0.7750**	**0.8073**	**0.8474**	**1.0026**	**0.6276**	**0.6217**	**0.5524**	**0.5434**
b0	1.0438	0.9081	0.9413	1.2008	1.3227	1.2376	0.9081	1.0407	1.0394	1.0383
b1	0.8840	0.9610	1.0171	1.1341	1.0362	1.1077	0.8840	0.8881	0.8744	0.8594
b3	0.6494	0.7524	0.6099	0.6398	0.7785	0.7711	0.6099	0.6501	0.6532	0.6610
b4	**0.7236**	0.6775	0.7236	0.8096	0.8269	0.7236	0.6775	0.7236	0.7236	0.7236
b5	**0.7922**	0.9654	0.8563	0.9493	0.8463	0.9157	0.7922	0.7922	0.7922	0.7922
b6	0.7476	0.7822	0.8611	1.0182	0.8577	0.7694	0.7476	0.7460	0.7454	0.7494
s0	0.6019	0.9305	0.9159	1.0439	1.0170	1.4421	0.6019	0.5425	0.3550	0.3513
s1	0.3860	0.4187	0.5222	0.3988	0.5416	0.8155	0.3860	0.3468	0.2579	0.2399
s2	0.4148	0.5314	0.6186	0.5248	0.6656	0.9462	0.4148	0.3727	0.2554	0.2234
s3	0.5475	0.7032	0.6678	0.7221	0.7118	1.0440	0.5475	0.5048	0.3804	0.3470
s4	0.4179	0.5616	0.6516	0.5173	0.7002	0.9947	0.4179	0.3727	0.2516	0.2620
s5	0.5435	0.7366	0.9145	0.7289	0.8639	1.2640	0.5435	0.4799	0.3011	0.2734

Optimization

- Solving the model as a MIP is much easier than solving the MIQP.
- Almost all of the submitted solution depends on first deriving the best MIP solution possible (i.e. minimizing the recurring load or minimizing the recurring + once-off load) and only then solving as an MIQP
- Gurobi 9.1.2 (laptop phase 1, UQ HPC phase 2)
- Various papers about “Predict+Optimize” problem but Phase 1 and leaderboard seem to indicate no close relationship between forecast result and cost. Complex problem, competition issues, limited time

NOW THAT WE HAVE A GOOD STARTING POINT …

- **Conservative** is just choosing the lowest recurring load and lowest recurring + once off load and evaluating cost using a naive or flat forecast. This was probably the winning approach for cost in Phase 1, as some competitors had winning results with no forecast, or a poor forecast, but seemed pointless to me as the organizers said quality of forecast should contribute to results in phase 2.
- **Forced discharge** forbids any charging in peak hours, and forces at least one of the two batteries to be discharging in every peak period.
- **No forced discharge** forbids any charging in peak hours, but the MIQP solver decides whether to discharge or do nothing in those hours.
- **Liberal** allows charging in peak, but the maximum of recurring + once off + charge effect for each period is limited to the maximum of recurring + once off load over all periods. This is to avoid nasty surprises when the solver thinks that a period has low underlying load and schedules a charge (due to a low price in that period) but then accidentally increases the maximum load over all periods, which can be very costly.
- **Very liberal** allows charging over peak and does not attempt to control the maximum of recurring + once off + charge effect. This would be the best approach if the forecast was perfect.

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

ESTIMATED COST

Only Large2/Large4 had the once-off load in, all activities, in peak.
The estimated cost is very different from the real cost.

Winning solution had almost all once-off activities included.

Case	Estimated Cost	Actual Cost
small0	26681	34166
small1	26233	33682
small2	26251	33235
small3	26452	33977
small4	26107	33462
large0	26265	32417
large1	26666	33842
large2	25389	32841
large3	26010	33149
large4	25849	33334
Total	**261906**	**335107**

		bean	esmailbeigi	evergi	akylas	sample
small	0	34166	34509	35676	37281	57941
	1	33682	33265		36862	50096
	2	33236	32428		34342	59924
	3	33977	33136		38344	46427
	4	33463	32490		35263	99669
large	0	33417	32643		34644	46404
	1	33842	33055		34949	78291
	2	32841	31712		36050	42501
	3	33149	32219		35389	55874
	4	33334	32903		39390	52230
	total	335107	328359		362515	589357

Summary

- Random forest – 4 models for each hour
- Use daily BOM solar data + ECMWF hourly data + temporal variables
- Train buildings and solar together in groups
- Thresholding two solar series

- Arrays approach with 0-1 Mixed Integer Program (MIP)
- First minimize recurring and recurring + once-off load, then solve MIQP
- "No forced discharge" approach chosen from 5 approaches

LESSONS

- Luck is an important factor
- The most effective methods may be absolutely *ad hoc*
- “AI competitions don’t produce useful models”
- Ranger package with multithreading is very useful
- Python is much better than R for Gurobi

How has this helped in other work?

- Phase 2 data now public – meaningful improvement is difficult
- **Project SHIELD** – data quality indicator using ECMWF data
- **Cybersecurity** of energy data – Springer chapter to appear
- **Forecasting at CEDI** – hot water, meter failure
- **E-scooters, e-bikes:** article in The Conversation, J of Transport Geography – demand forecasting works in same way
- What *are* the best solar/load predictor variables in an Australian context?

Thank you

Contact information

R.Bean1@uq.edu.au