Forecasting Hydro-Power in Brazil

Mentors: Ivan Marin, Tanmay Raj

Bootcamp Coordinators: Daniel Spirn, Thomas Höft

Ana Chavez Caliz (PSU), Jürgen Kritschgau (ISU), Francisco Martinez (OSU), Avishek Mukherjee (UDel), Smita Praharaj (UMC), Cameron Thieme (UMN), Jennifer Zhu (TA&M)

Cargill in Brazil

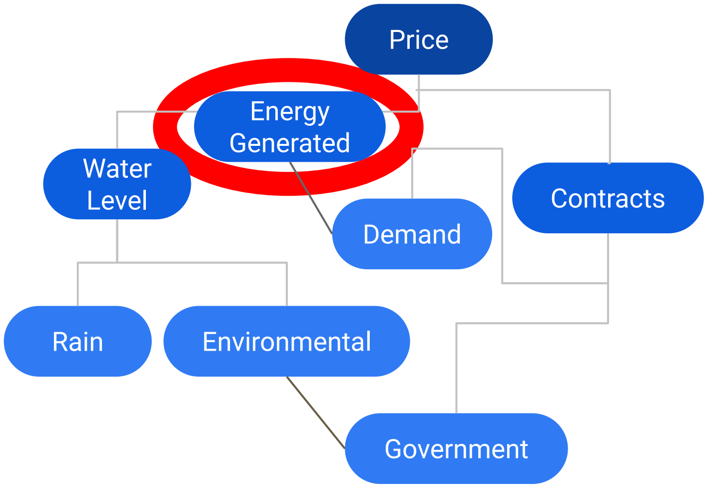
Cargill's soy facilities and sourcing areas in Brazil

- 122.6 million tons of soy production*
- 15,000 suppliers selling soy to Cargill*
- Soy processing plants can be powered by Cargill or 3rd party.

Problem: Should Cargill buy electricity or produce it?

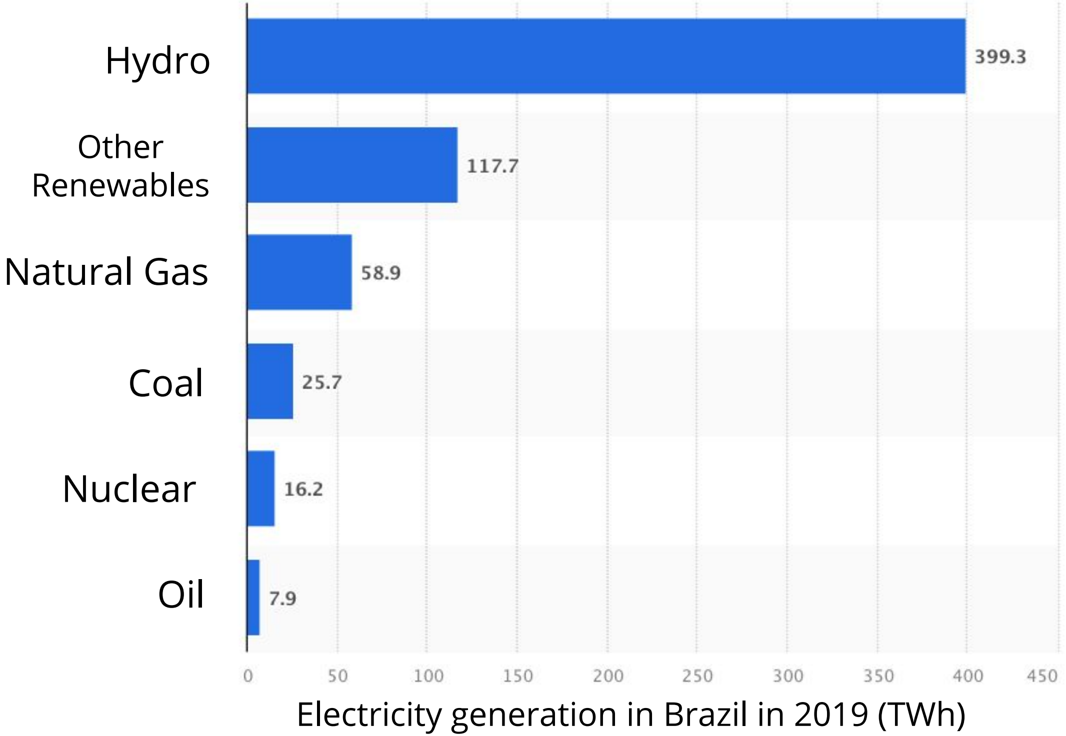
*2020 Cargill report

Our Role



- **Ideal Goal: Predict energy price.**
- **Obstacle:**
  - Price depends on production, demand and complex government regulations.
- **Alternative: Predict energy generated.**
- **Advantages:**
  - This prediction can be used to predict energy price in the future.
  - Depend on less variables.
  - More accessible data.

Hydroelectricity in Brazil



Hydropower accounts for 70% of Brazil's power (~16% of the world's power is hydro!).

Goal: Predict hydroelectricity generation (~12 months in advance)

Timeline

Data Collection

Compiles:

- Hydroelectric system information.
- Weather information.

At different geographic levels:

- Local.
- Regional.

Translated to English.

Distinguish significant variables

- Energy demand.
- Rainfall.
- Influent flow.

Determine time frame and scale

- Monthly analysis.
- Regional analysis.

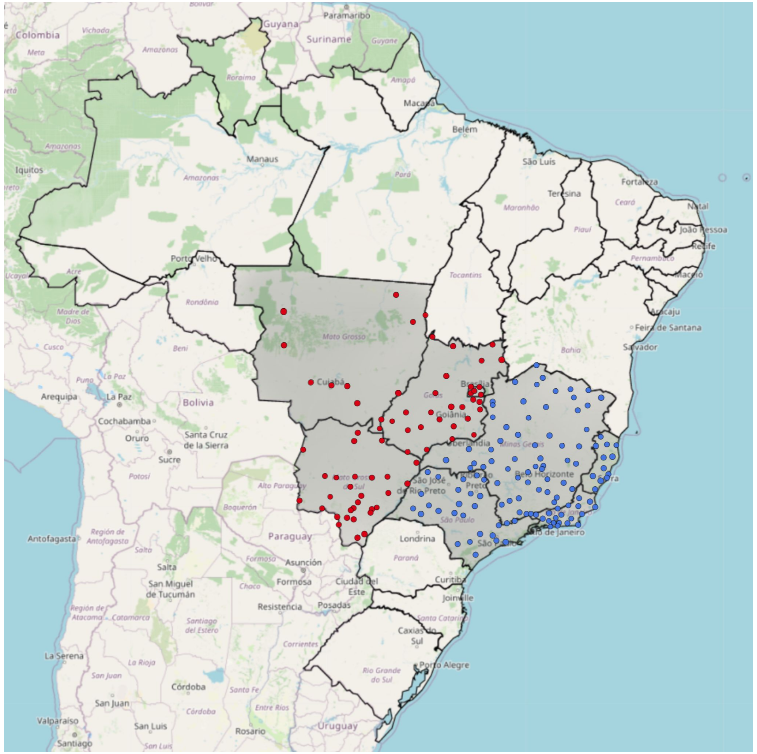
Prediction model

Creation of a model that forecast energy generated:

- Univariate time series models.
- Incorporating exogenous regressors.
- Recurrent Neural Nets.

Data Collection Process

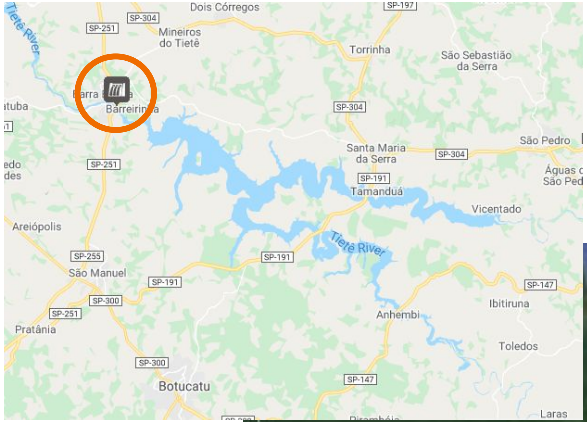
- Dam and power grid data obtained from the ONS
  - Daily, weekly, monthly scale depending on variable.
  - Generally complete, with few missing entries.
- Weather data from mixed sources
  - Aggregated to basin, state, or region.
  - Daily, weekly, monthly scale.
  - Very patchy data.
  - Averaged by dropping “NA”s.



Weather Stations (SE/CW region)

Understanding Energy Generation at a Single Dam

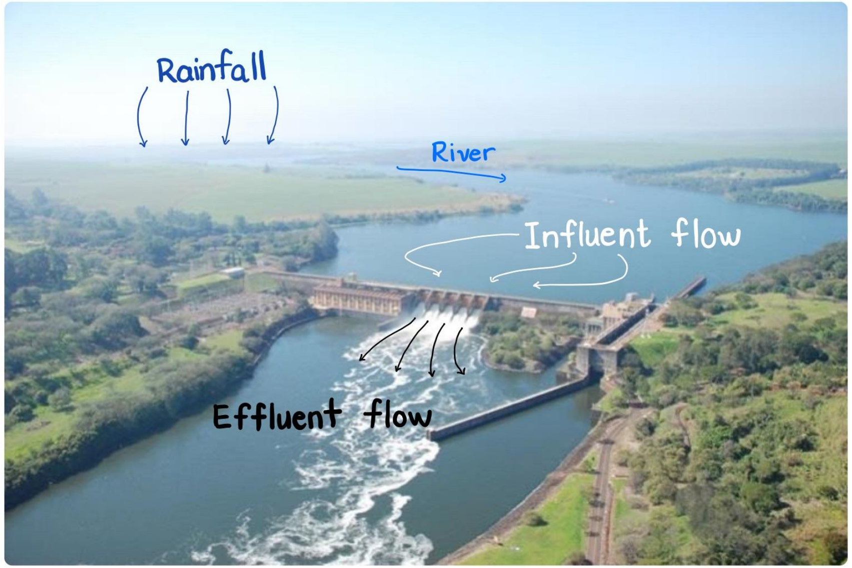
Barra Bonita Hydroelectric Power Plant



- **Location:**
  - State: São Paulo.
  - River: Tietê.

- **Simple Dam**
  - High water level.
  - Independent: no dams upstream.
  - Generation capacity: 140 MW.
  - Big reservoir: stretches upto 150 kms.

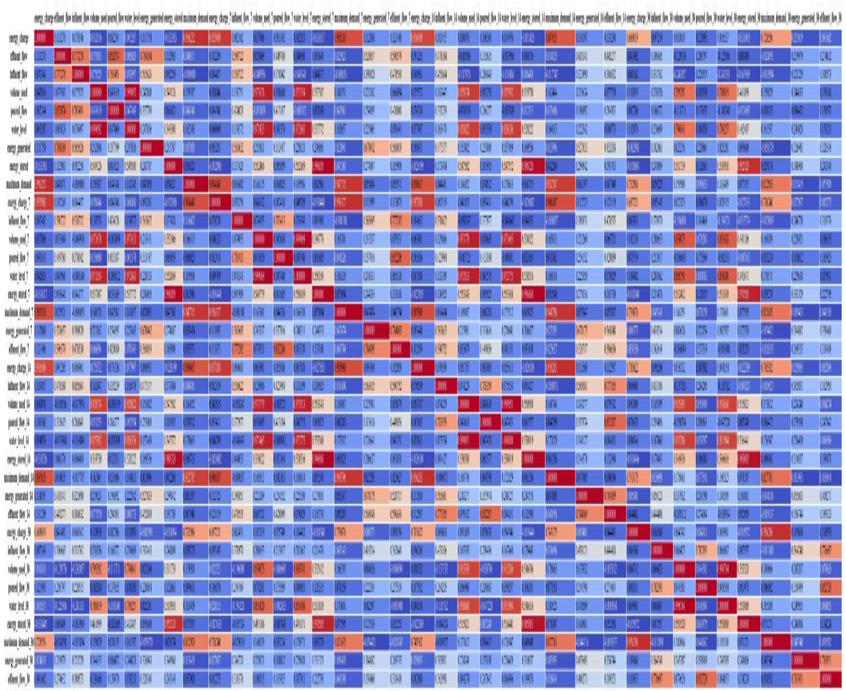
Significant Variables at Dam Level



Forecast: Energy generated

- **Energy demand:** for the subsystem (MWh/h)
- **Rainfall:** by basin (mm)
- **Influent flow:** Water that the reservoir receives from rainfall and other natural resources (m^3/s)
- **Water level:** Height of the water in the reservoir measured above the sea level (m)
- **Lag of variables:** Incorporating information from the past.

Correlation of significant variables

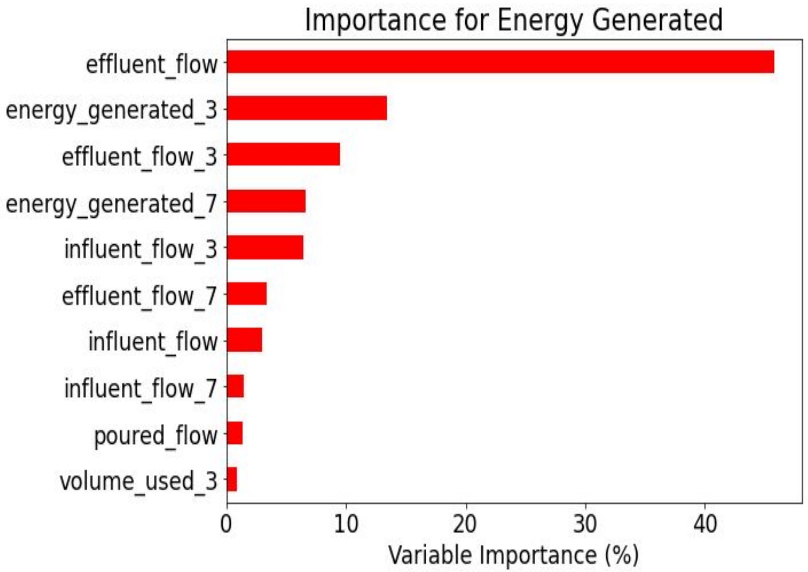


Correlation matrix of all variables at B. Bonita lagged 0, 7, 14, 30 days

- Only strong correlations are trivial (eg: turbine flow vs energy generated).
- Models at the dam level were not accurate.

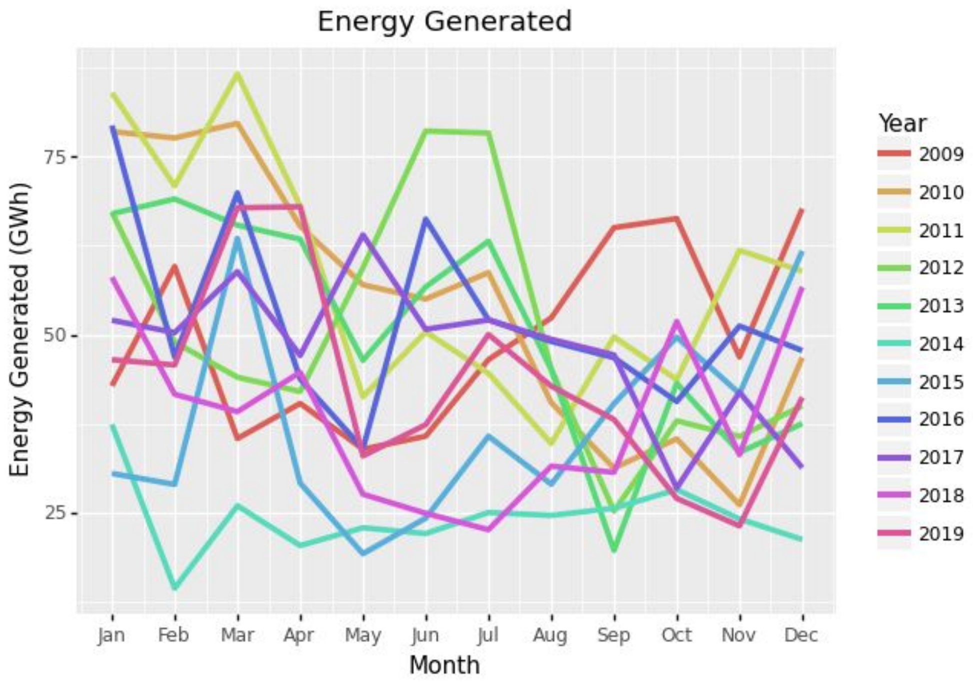
	energy_charge	usable_volume	water_level	energy_stored	maximum_demand
energy_charge	1.000000	0.008105	0.010377	-0.081993	0.958172
usable_volume	0.008105	1.000000	0.999120	0.556795	0.123996
water_level	0.010377	0.999120	1.000000	0.557616	0.125787
energy_stored	-0.081993	0.556795	0.557616	1.000000	0.006074
maximum_demand	0.958172	0.123996	0.125787	0.006074	1.000000

Finding Significant Variables



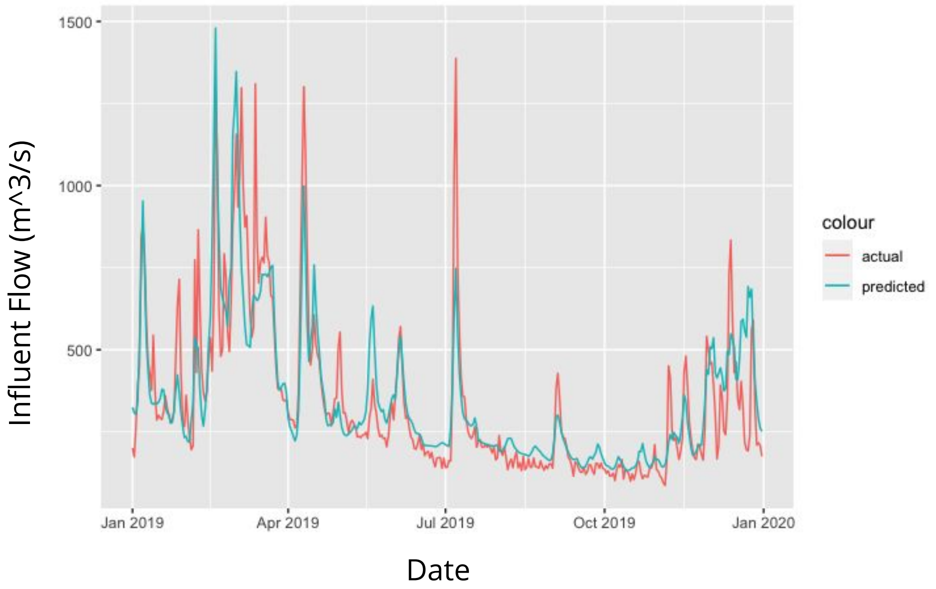
- Trained a random forest with 500 trees.
- Trees have depth 7.
- Importance of a variable is a weighted sum of error reduction.
- Most Important:
    - Effluent flow
    - Energy generated (lag 3)
    - Effluent flow (lag 3)
    - Energy generated (lag 7)

Energy generated at Barra Bonita



- Energy generated doesn't exhibit a clear seasonal behavior.
- Other variables (energy demand, rainfall, influent flow, water level) are seasonal.

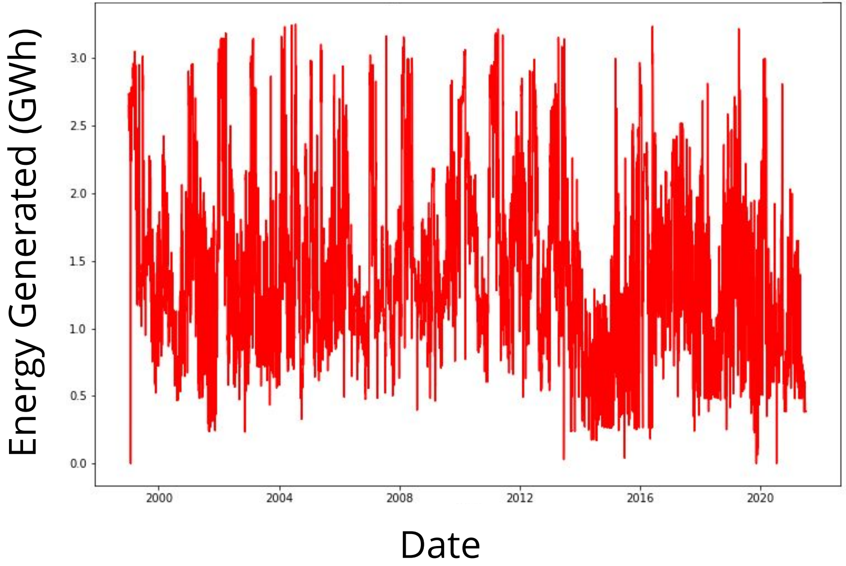
Influent flow at Barra Bonita



- Trained on 2007 - 2019 data
- Quantifies Relationship: **Not a Prediction**
- Clear relation between rain and influent flow.
- Most important lags are from 2 and 3 days ago.
- MAE about 23% of the mean influent flow.

log(influent_flow) ~ (2 weeks of lagged rain) + water_level + month

Conclusions for Energy Generated at Barra Bonita

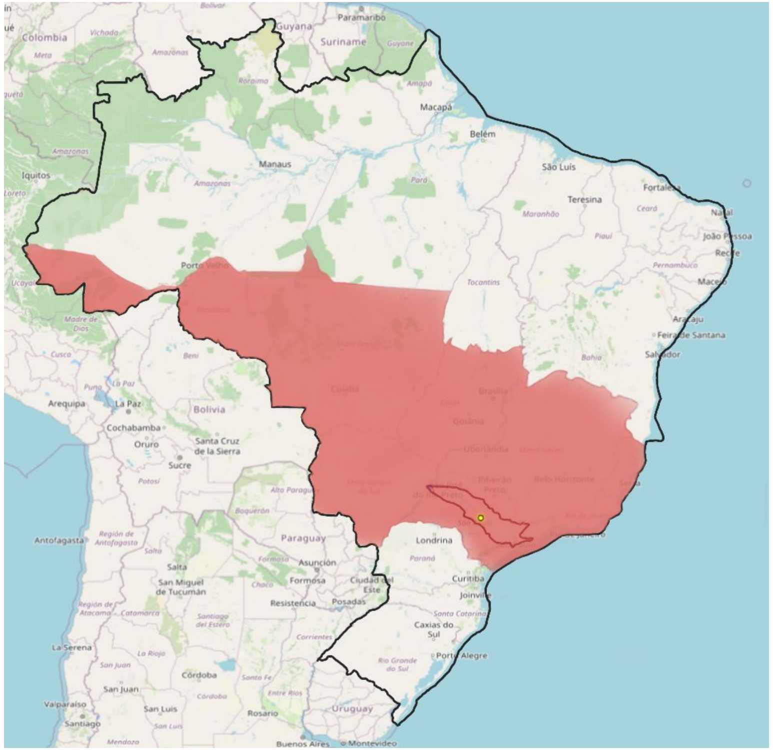


- Correlations between variables at the dam level are weak at best.
- Best models we made are very inaccurate:
    - MAPE around **50%** for all models.
- No exploitable seasonalities at the Barra Bonita level.
- Some success in quantifying the relationship between rainfall and influent flow

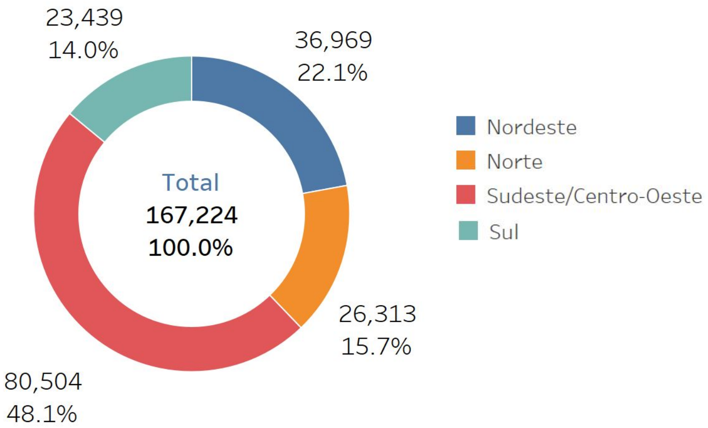
Rescaling the Problem

(Look at a bigger region)

Southeast/Central West Subsystem



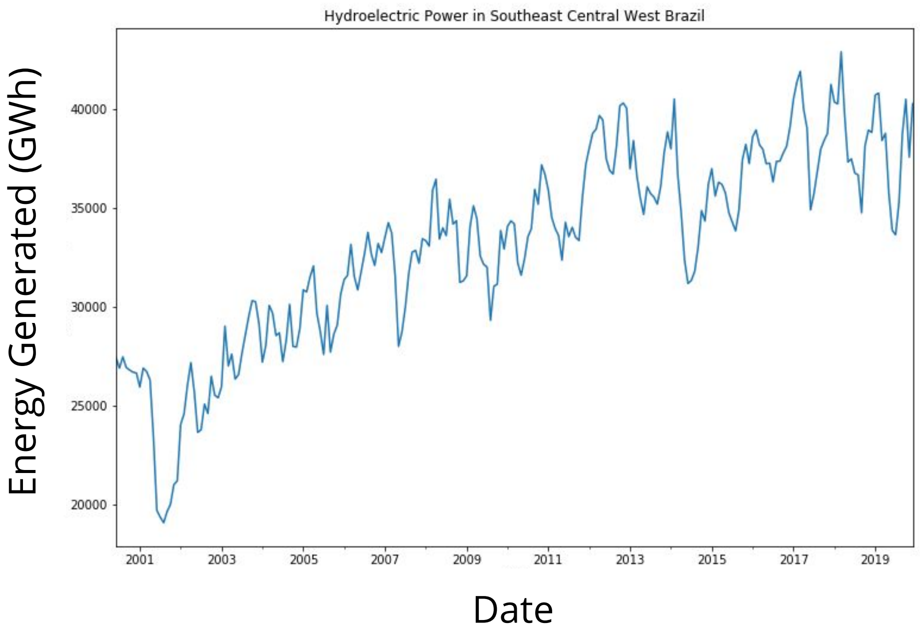
Hydroelectric Capacity by Region*



- 8 main basins.
- 20 main reservoirs.
- 101 Power plants.
    - Including Itaipu, the second largest Hydroelectric power station in the world.

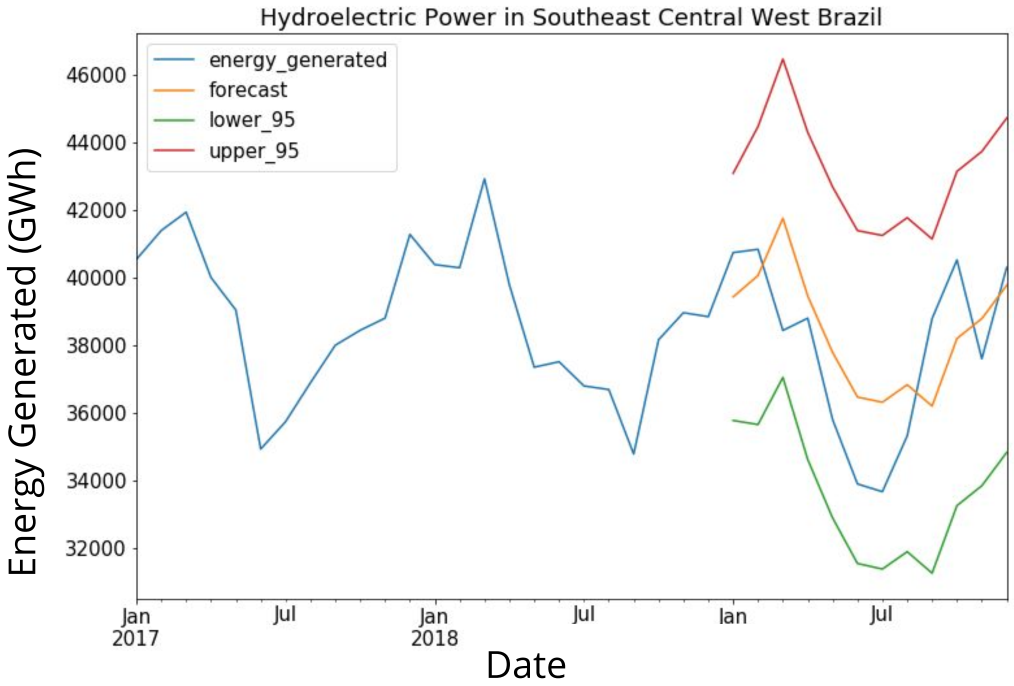
* graph from ONS

Energy Generated at Subsystem Level



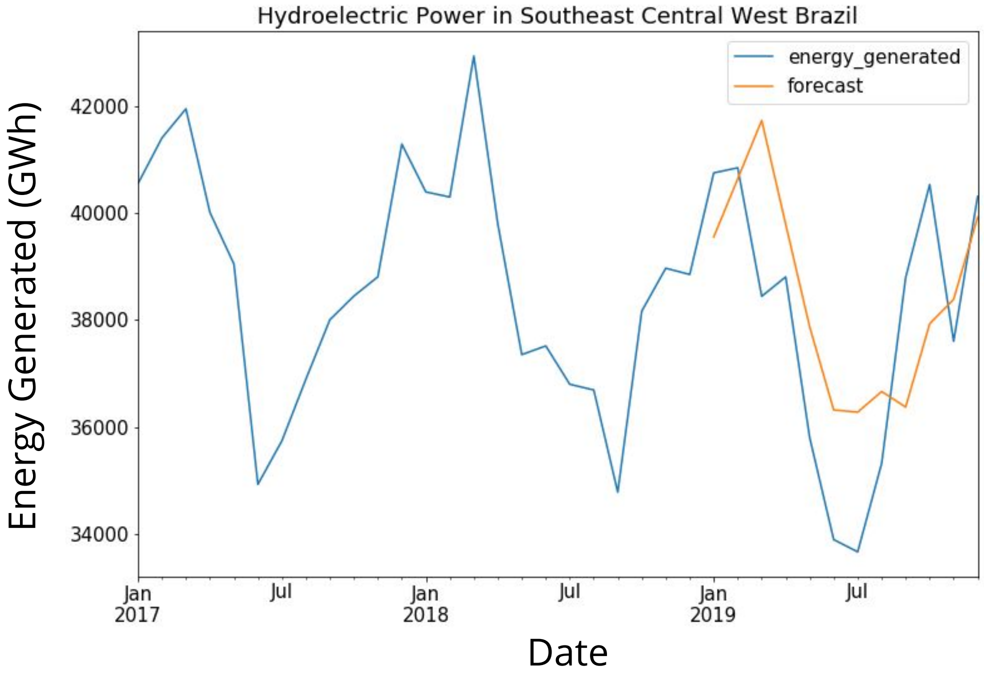
- Total energy generation in Southeast Central West Brazil.
- Monthly data.
- Forecasted using standard time series methods:
  - Naive, Seasonal Naive, SES, Holt Linear, Holt Exponential, Holt-Winters, Damped versions of the exponential smoothing models.
- SARIMA chosen for final predictions
  - Slightly improved accuracy.
  - Ability to incorporate exogenous regressors.

Without Exogenous Variables



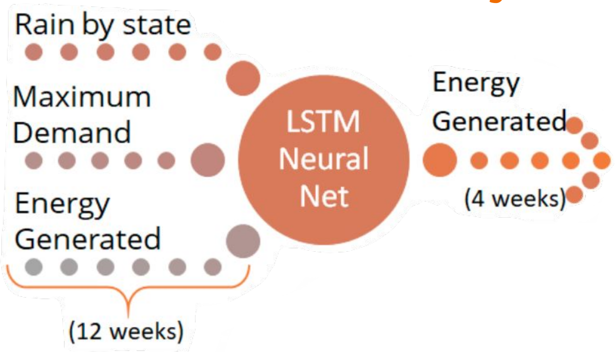
- CV MAPE between **3.3%** and **4.1%** for predictions one to twelve months in advance.
- Errors do not depend on the month.
- Training data larger than shown.

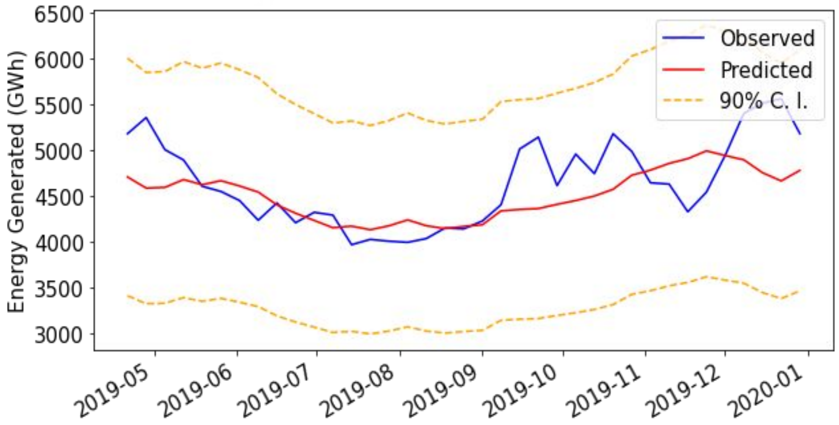
With Exogenous Variable



- **Mean of influent flow** at 15 largest available dams as exogenous variable.
- Two layers of SARIMA: predict influent flow, use that in energy generation prediction.
- CV MAPE between **3.1%** and **4%** for predictions one to twelve months in advance.
- Errors do not depend on the month.
- Other exogenous regressors had similar results.

Neural Net Weekly



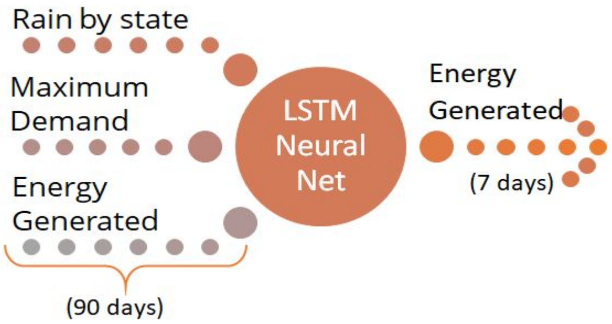


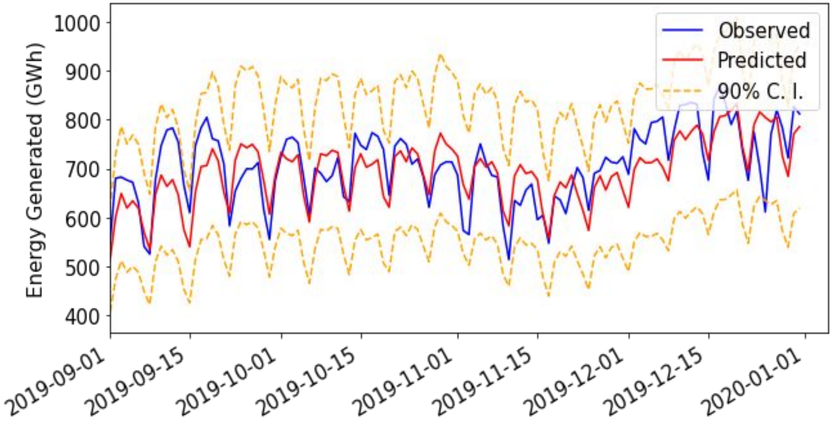
- 1026 weeks of data, separated into train, validation and test sets.
- 11 input variables.

- Mean absolute error of **6.1%**
- Naive prediction of last known week, MAPE of **8.9%**
- Architecture:
  - Two LSTM dense layers, of 15 nodes each.
  - Dense layer of 4 nodes.

- Less accurate than SARIMA.
- Needs accurate rain.

Neural Net Daily





- 7178 days of data, separated into train, validation and test sets.
- 11 input variables.

- Mean absolute error of **5.4%**
- Naive prediction of last known week, MAPE of **6.1%**
- Architecture:
  - LSTM dense layer of 10 nodes.
  - Dense layer of 7 nodes.

- Less accurate than SARIMA.
- Needs accurate rain.

Future Directions: Determine Optimal Scale

- Examine different time and space scales
  - Predicting daily output at a small dam is hard
  - Predicting monthly output in the region is doable
  - What can be said about levels between this?
    - Weekly data
    - Basin level, dam level
- Inter-dam relationships
  - Dams located along the same river may influence each other’s power generation.

Thank you for your attention!

References

- ONS - National Electricity System Operator - http://www.ons.org.br/
- INMET - National Institute of Meteorology - https://portal.inmet.gov.br/dadoshistoricos
- CPTEC - Weather Prevision Center and Climate Studies - https://bacias.cptec.inpe.br/#!