[SOLARISE]

Enercoop short term forecast algorithms

09 march 2021
Pierre Gouëllo

100% COOPÉRATIF
FOURNISSEUR D'ÉLECTRICITÉ
100% RENOUVELABLE

enercoop
L'énergie militante

Plan

1. Day Ahead Load Forecast (1h time step)
2. Limiting factors on accuracy
3. Production forecasts

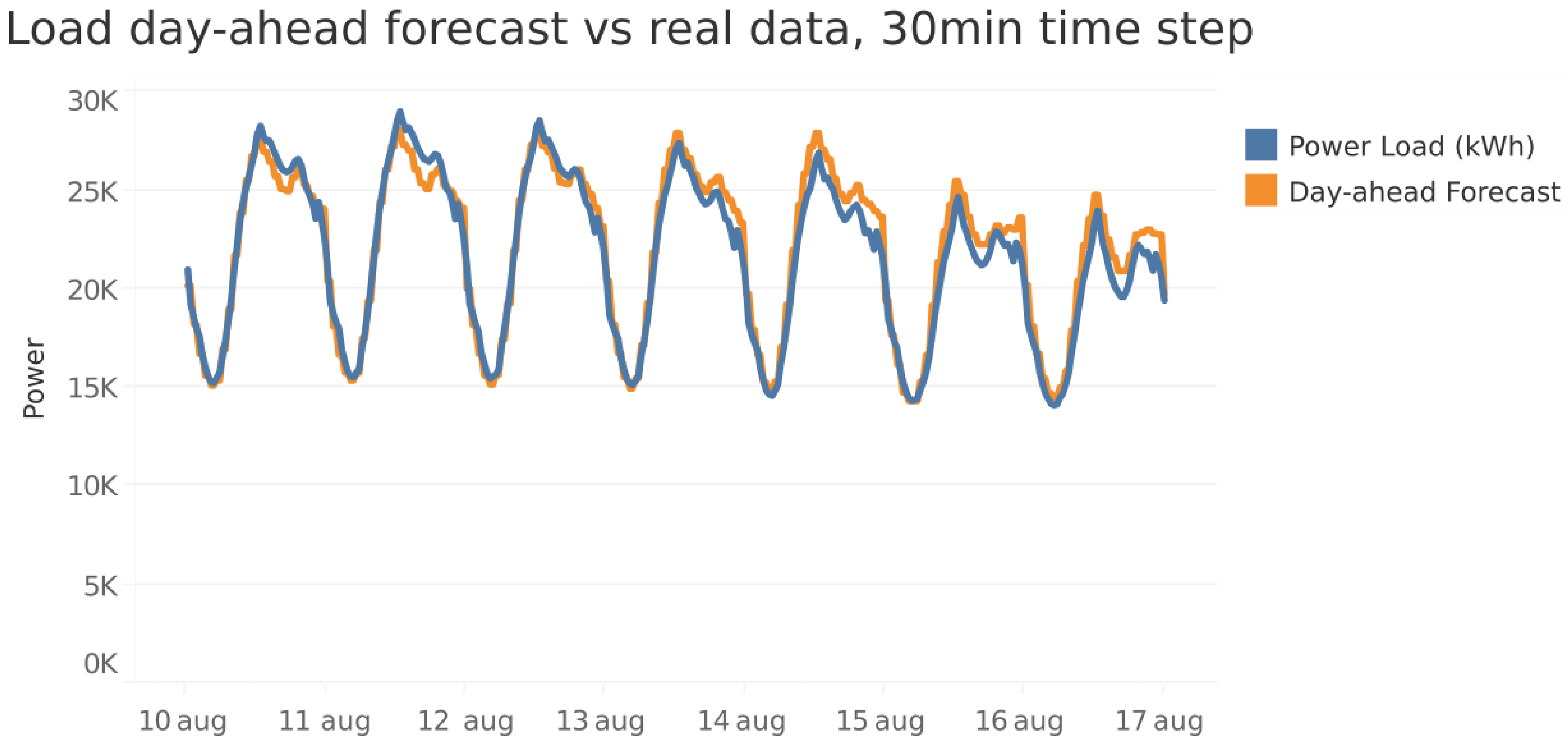
Part 1

Day Ahead Load Forecast
(1h time step)

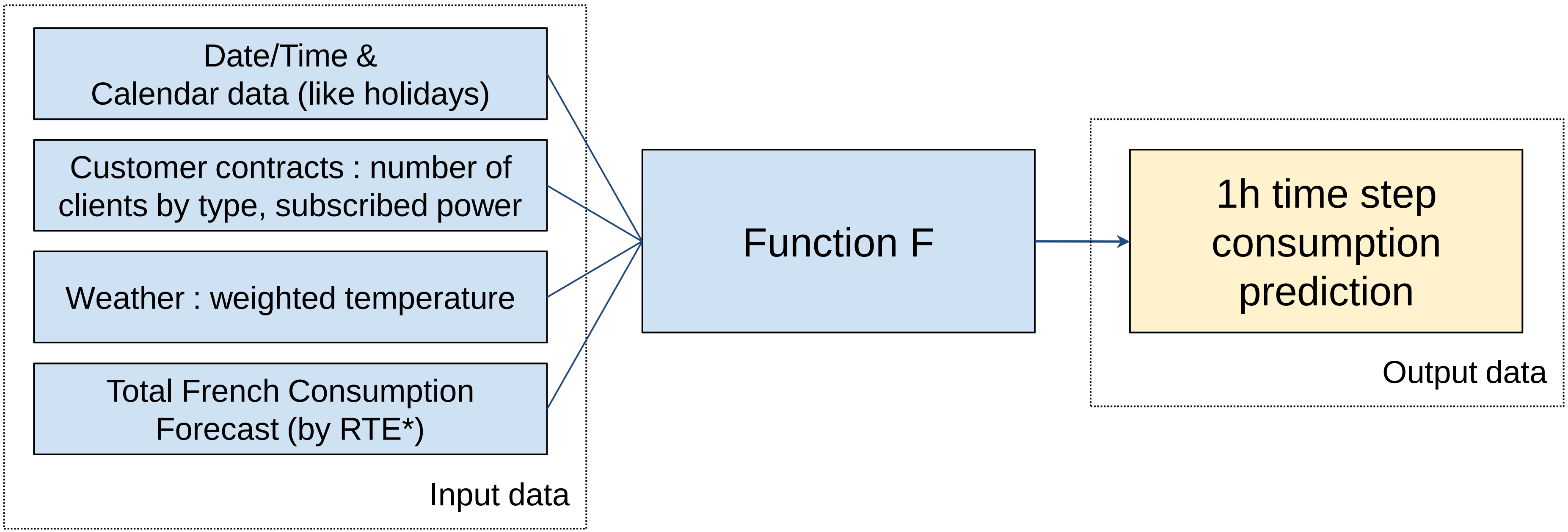
100% COOPÉRATIF
FOURNISSEUR D'ÉLECTRICITÉ
100% RENOUVELABLE

enercoop
L'énergie militante

Day-ahead load forecast : result



Day-ahead load forecast : predict

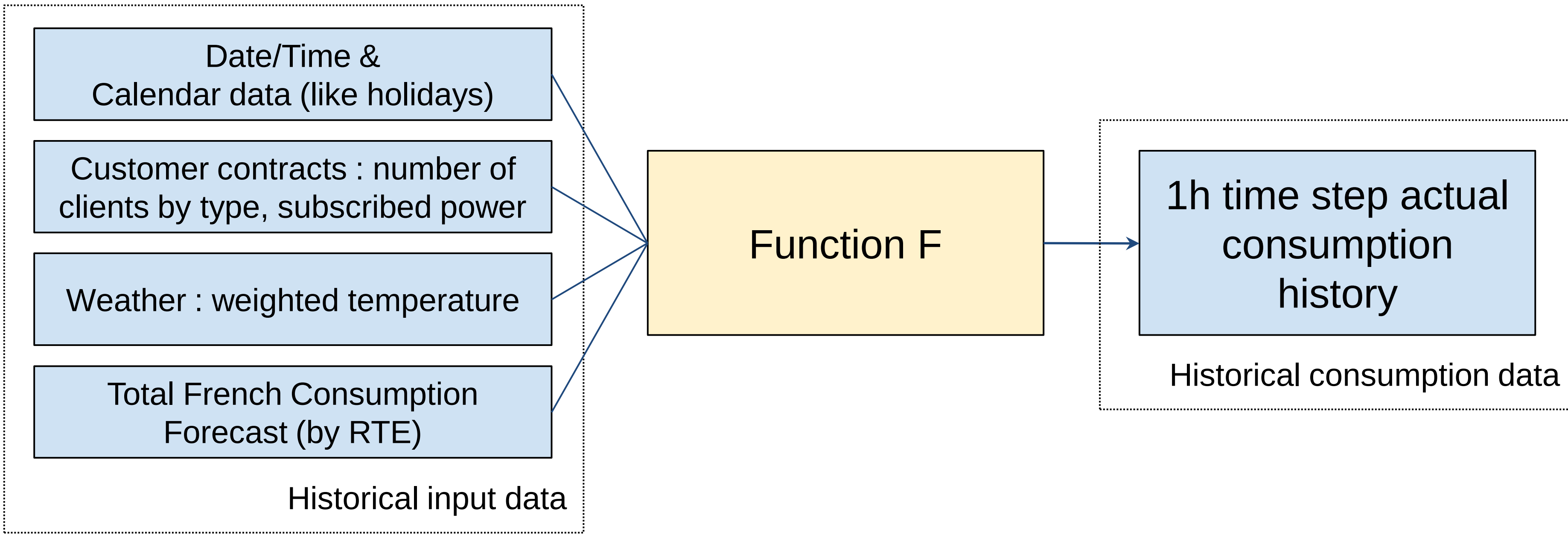


Predict mode :

- input data is **predicted** for each hour in the next 7 days
- function F is applied for each hour in the next 7 days on this predicted input
- prediction is run daily
- it generates automatic day-ahead transactions on the SPOT market to even out Enercoop's energy balance (we can count in € the cost of imprecisions)
- Current precision : usually around 2-3% (our consumption ~= 50MW)
  - but sometimes around 6-8%

* RTE : French TSO (transmission system operator)

Day-ahead load forecast : train



Train mode :

- function F is "trained" using **machine learning** algorithms on **historical data**
- recent data is required for the function to adapt to current consumption trends
- a statistically-significant volume of historical "training" data is required for F to be precise
- F is re-trained every week automatically

Key aspects

- Input data is predicted
- Quality of load forecast = quality of input data
- Quality matters for both day-ahead data and historical training data

Part 1

Details

100% COOPÉRATIF
FOURNISSEUR D'ÉLECTRICITÉ
100% RENOUVELABLE

enercoop
L'énergie militante

Day-ahead load forecast: historical data

load historical data
(hourly step)

- Load historical data for each synthetic load profile + smart-metered customers
- Usually we have to combine multiple sources such as old csv files, data coming from former BRPs*, etc...
- For some profiles we have a low volume or we lack of historical data => we choose to aggregate them with other profiles
- Delays : the definitive load value measured by RTE comes after 14 months.

* BRP = balance responsible entity

Day-ahead load forecast: Calendar data

Date/Hour &
calendar data

- Usually it is easy to obtain holidays, weekdays, etc… from officials sources (government, DSO, …) once and for all, however some unexpected events always happen (ex : covid lockdown which looks like a holiday in the point of view of the algorithms)
- “Feature engineering” is important
- Load periodicity : hour, day, week & weekends, seasonal
  - => need a lot of historical data : a few years
- Typical special days
  - school holidays, summer holidays, national holidays, lockdown

Day-ahead load forecast: Portfolio

Portfolio : number of customers and kVA per profile + smart meters

- It is extremely important to know the exact portfolio everyday to make an accurate forecast
- The sources of data must be the same fore algorithm training and forecasting. Typically, in can come from the ERP/CRM
- Delays exist in information systems so it is hard to have the precise portfolio
- In our case, we still have to do a few additions manually : for example anticipating the arrival and departure of a big professional customer

Day-ahead load forecast: Weather model

Weather : temperature

Actually the weather forecast is a model of its own

- We forecast a unique temperature for each hour : the average metropolitan French temperature (weighted average according to population density)
- “Météo France” forecasting is processed via a forecasting algorithm.

Meteo France weather forecast (we keep about 30 cities)

Population of these cities

Weighted average → Function T1 → Temperature

- Function T1 is trained with:
  - Input: historical data form Météo France
  - Output: historical data of the weighted average measured by meteorological stations (SYNOP)
- We also use a T2 function that exploits data from the american model GFS. It is less accurate for France but more robust. We typically use it when the Meteo France data are not available.

Day-ahead load forecast: TSO Data

Total French Load (RTE)

- RTE (our TSO) provides hourly forecasts of the national load for the next 8-10 days.
- We retrieve them from their API and use them as input in our forecast
- To remain robust in case their prevision is amiss:
  - We train part of our function F with RTE data and part of it without.
  - Then we get a weight for each part of F based on recent data.
  - Thus we automatically optimize whether or not we use their data as input

Day-ahead load forecast: The function F

Function F

F is just a static function (in prediction mode) that takes X variables as input and outputs a variable Y.

- It is divided in two functions:
  - F1 = forecast per kVA* for synthetic load profiles x kVA expected
  - F2 = forecast of smart-metered customers
- Each function (F1, F2) is trained in 2 steps :
  - train of ~10 classical models** on long historical datas (5-10 years)
    - 5 models trained with TSO data and 5 without
  - global training to combine the 10 models according to recent historical data (1 month)
    - follows recent evolutions, weights the different models
- Accuracy as of today : about 2% error with 6-8% rather often.

* kVA : power unit to describe subscribed power

** classical models : GAM, XGBOOST, GBM, Random Forest, Neural network, etc.

Part 2

Limiting factors on accuracy

100% COOPÉRATIF
FOURNISSEUR D'ÉLECTRICITÉ
100% RENOUVELABLE

enercoop
L'énergie militante

Limiting factors on accuracy

Algorithmic quality can only get us to a certain point, the rest is about input data & historical input data :

- calendar data : easy
- weather forecast : need to use robust APIs
  - we use Meteo France data with a fallback on GFS data.
- total french load forecast (RTE) : connected to RTE's API
  - no SLA on their API, so we actually have a prediction pipeline without this data
- **day-ahead customer contract data** is the most difficult :
  - delays exist between information systems of Enercoop and our TSO
  - separate fixed-term and indefinite contracts
  - fixed-term contracts are usually large enterprises and can imply big leaps in total subscribed power
    - must connect to ERP or any up-to-date data-source to detect them ahead
    - usually we know the start date and end date of these contracts
  - indefinite contracts are usually residential or small professional customers
    - they can join or leave without warning
    - we use our recent commercial trends to forecast this

Part 3

Production forecasts

Production forecasts

Our R&D started experimenting short-term production forecasts. We will make it operational in the next 1 or 2 years.

The algorithm will predict an **aggregated production** or a **production per plant** using historical data. The algorithm is different for each technology. In 2020, we have about 300-400 power plants to forecast. Input data is :

- calendar data : mostly to detect seasonal periodicity
- weather forecast data (required for every plant) : solar/wind/hydro data
- recent measured weather data is also used
- **the plant portfolio and their state** is the hardest to gather :
  - peak-power / technology / location of each plant is easy
  - an up-to-date data-source of maintenance events is required
    - maintenance events can be full or partial shutdowns (like 1 out of 3 wind turbines)
  - the history of production but also the history of maintenance events (and of peak-power if it changed over time) for each plant is required for training
- for plants with little historical data, we will rely on historical data from other similar plants, and also construct specific feature engineering.

Thanks !

Sorry for the verbose slides !

Pierre Gouëllo