Generation of a future-proof hydro inflow dataset for power system studies

Simone Biondi [1], Nicholas Fabio Bonfanti [1],
Sara Maffioli [2], Manuele Aufiero [2], David Radu [3]

1. HelioSwitch, Milano, ITALY, Italy
2. Milano Multiphysics, Milano, Italy
3. ENTSO-E, Bruxelles, Belgium

spin off®
POLITECNICO DI MILANO

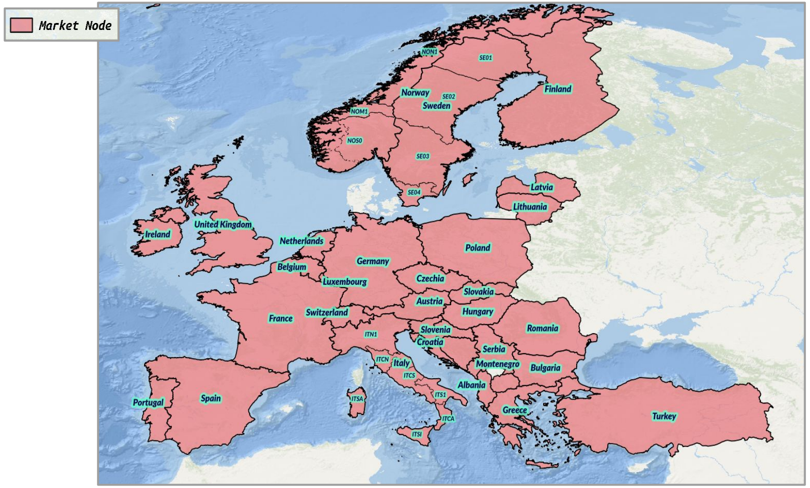
HelioSwitch

PoliHub
Innovation Park
Startup Accelerator

KEY DEVELOPMENT OBJECTIVES

The key points driving this development of the Hydro Inflow Dataset for Power System are the following:

- Have a **unified framework** to estimate historical and projected inflows for **each market node**.
- Requiring only minimal input data, i.e. the **historical production**, **river discharge** data from ECMWF, **power plant location** (optional).
- **Not requiring** power plant network topology, measured natural inflows, reservoir levels.



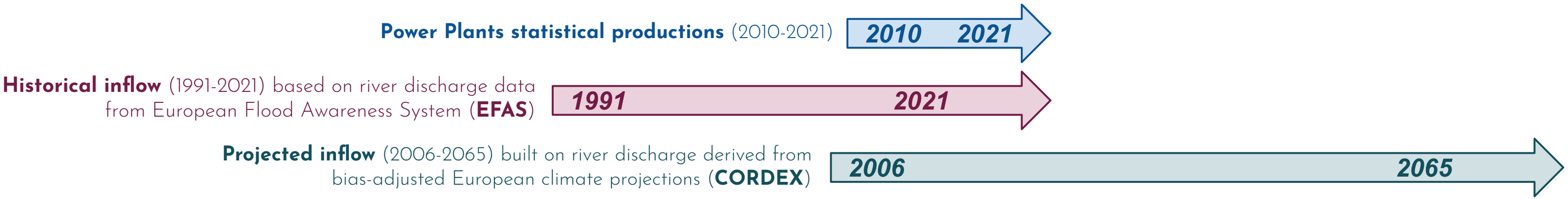
A key hypothesis of the model is that the **production is a good proxy of the inflows**, on a **daily basis** for RoR and Pondage, and on a **weekly basis** for Reservoir and Pumping.

The model will thus focus on the **transfer function between river discharge and production**, leaving the inflow estimation as an aggregation in time from the productions.

PROCESS OVERVIEW

The designed process is structured as follows:

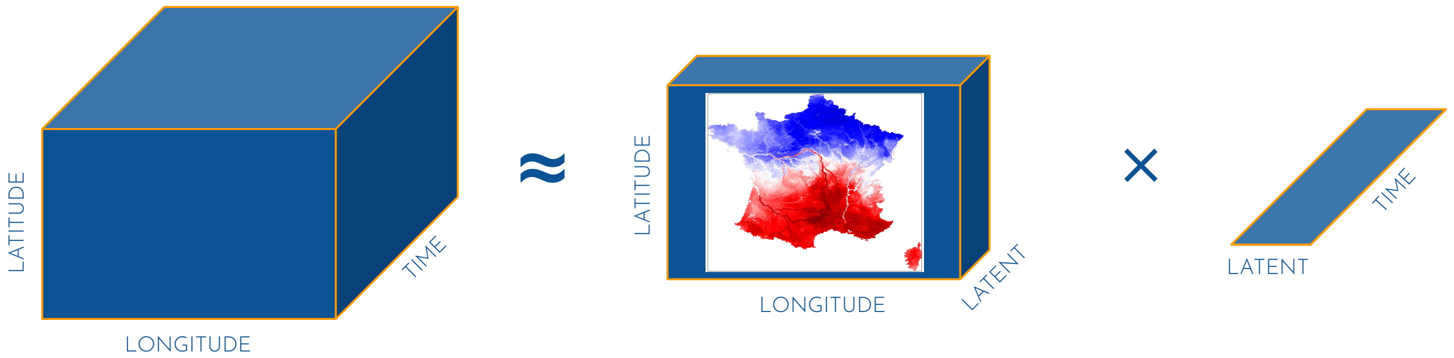
- Definition of the "**EFAS** river discharge → production → inflows" **transfer function**
  - **Step 1:** river discharge **dimensionality reduction**
  - **Step 2: non-linearity** from river discharge latent space to production data
  - **Step 3:** transfer function quality check and **evaluation**
  - **Step 4:** resampling and **interpretation as inflows**
- **Back-casting** with **EFAS** and projection with **CORDEX** river discharge datasets



- Handling power plant **mis-categorization**, missing metadata, and **anomalous behaviour**
  - **Frequency analysis** of the power plants generation time-series
  - Example: analysis of **reservoir** PP **upstream of RoR** PP.

RIVER DISCHARGE LATENT SPACE

The input river discharge data is **pre-processed with a PCA**, using the geographical bases to gain a physical insight, and the time-series coefficients as a pre-processed inputs for the downstream model.



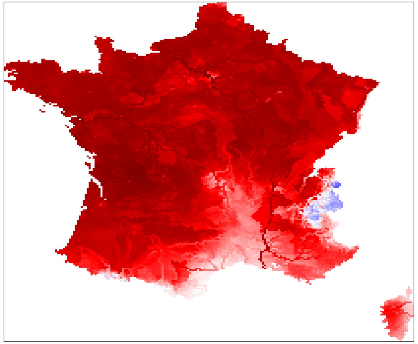
Full river discharge data.

Few **lat-lon maps** with **correlated** river discharge dynamics.

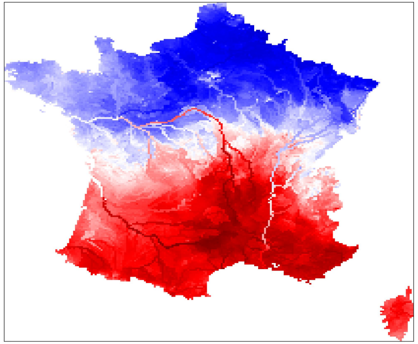
Few **time-series** describing the contribution in time of the lat-lon bases.

RIVER DISCHARGE LATENT SPACE

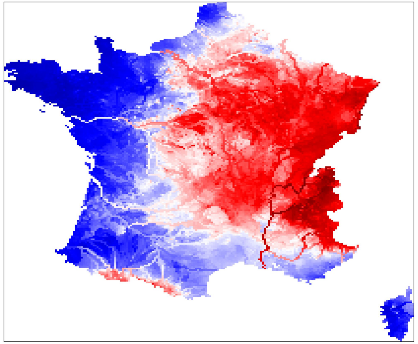
The **PCA identifies "physical" dynamics** in the river discharge data, for example in France:



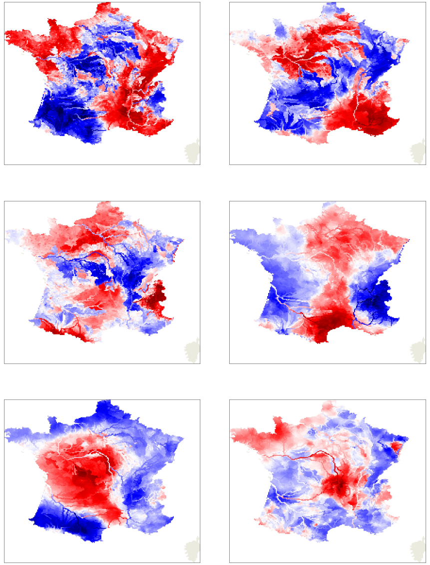
The first base is generally associated with the **average value**, and has **little spatial information**.



The second base highlights the **north-south differences**, and the correlation of the Loire river with south dynamics.

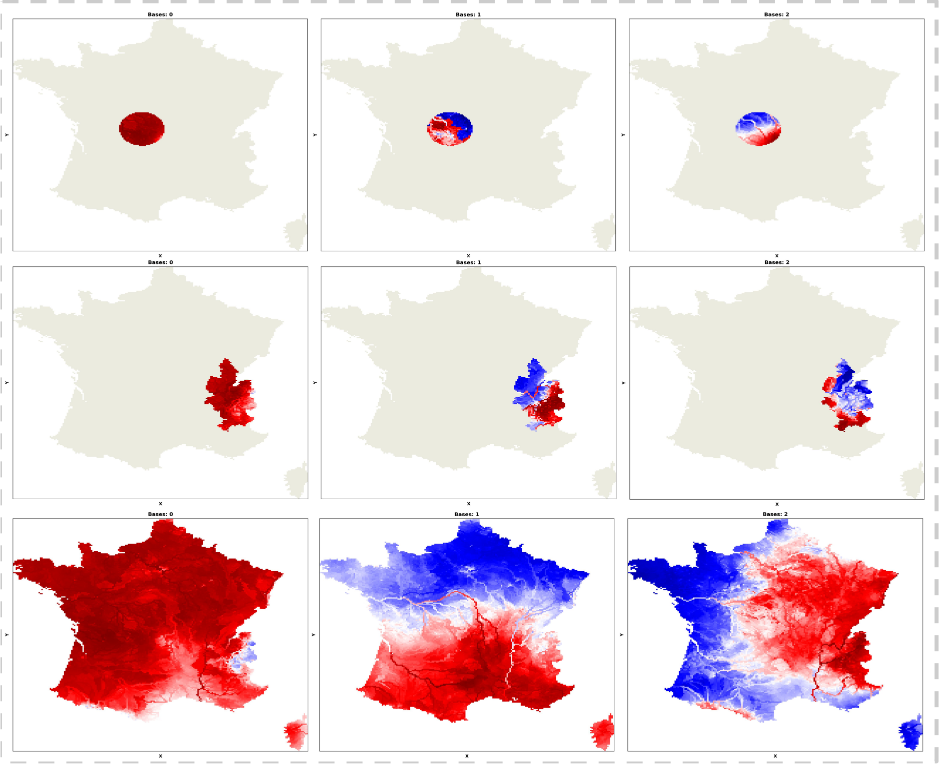


The third base shows a **inland-coastal differences**, with the Rhone river correlated with the inland dynamics.



RIVER DISCHARGE LATENT SPACE

If a power plant is **geolocated**, the **PCA** is only computed in its **proximity**.

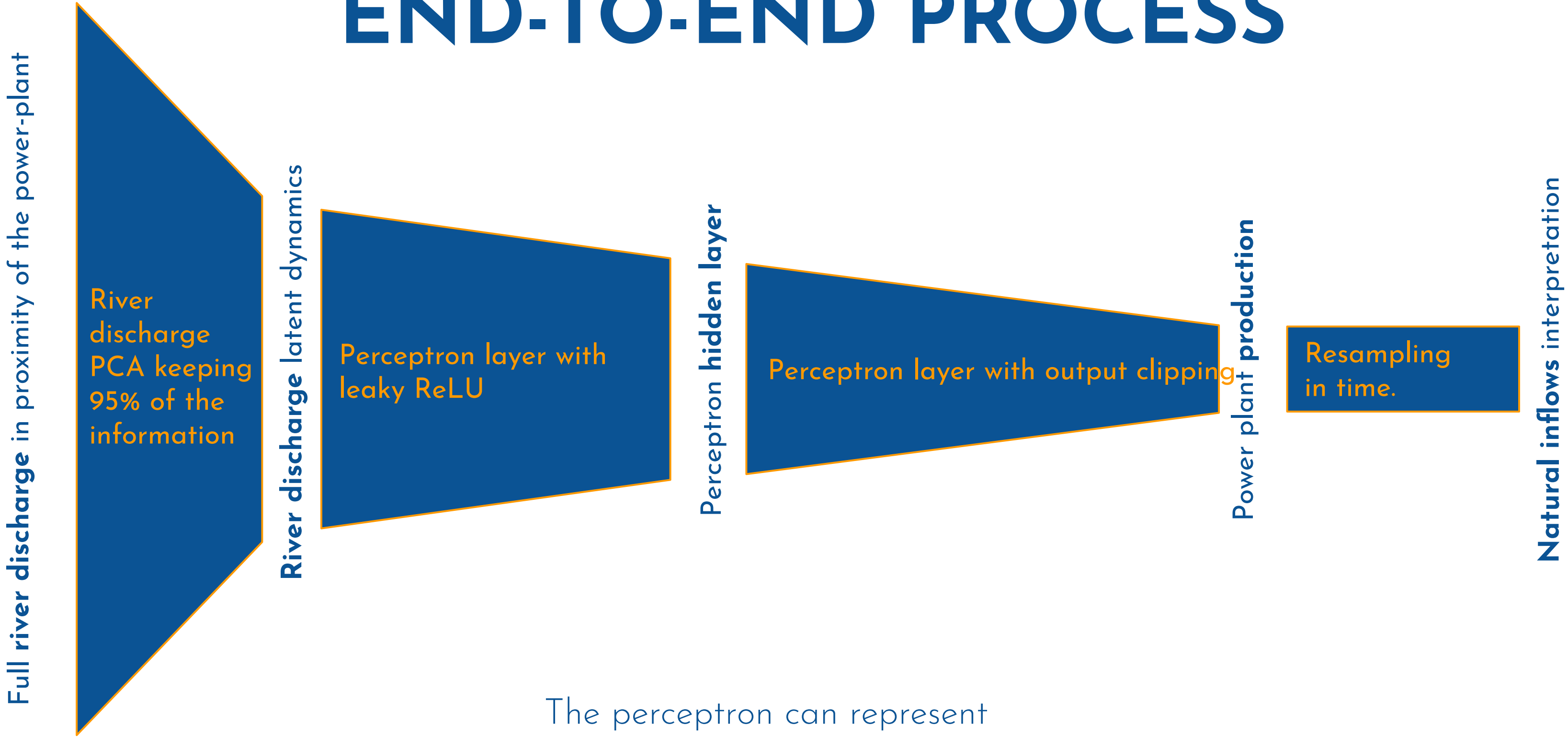


Known power plant **lat-lon** location

Known power plant **PECD** zone

Known power plant **country**

END-TO-END PROCESS

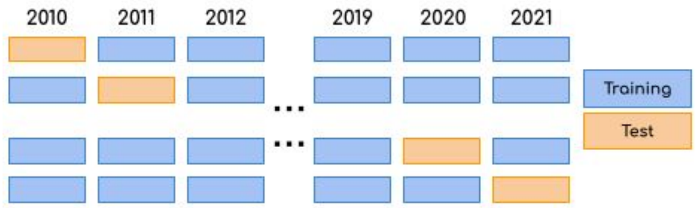


The **PCA** has a explainable and **regularizing** effect, capturing the physical dynamics around the PP.

The perceptron can represent both the weakly **nonlinear** relation between river discharge and production, and the strong nonlinearity of the **saturation.**

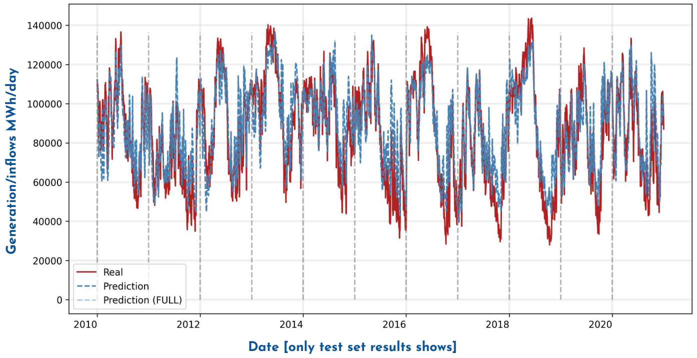
Resampling data over a period T allows to exclude reservoir dynamics with dynamics <T and work with the production ≈ inflows.

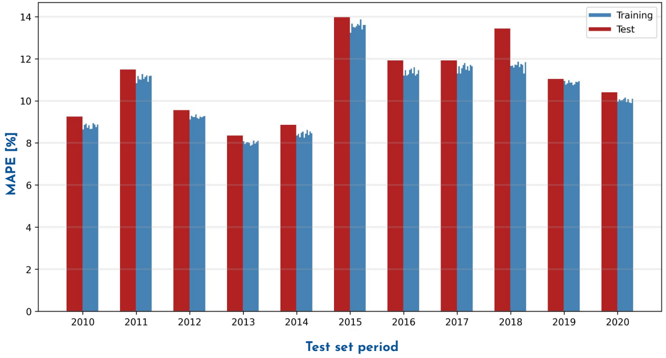
MODEL VALIDATION



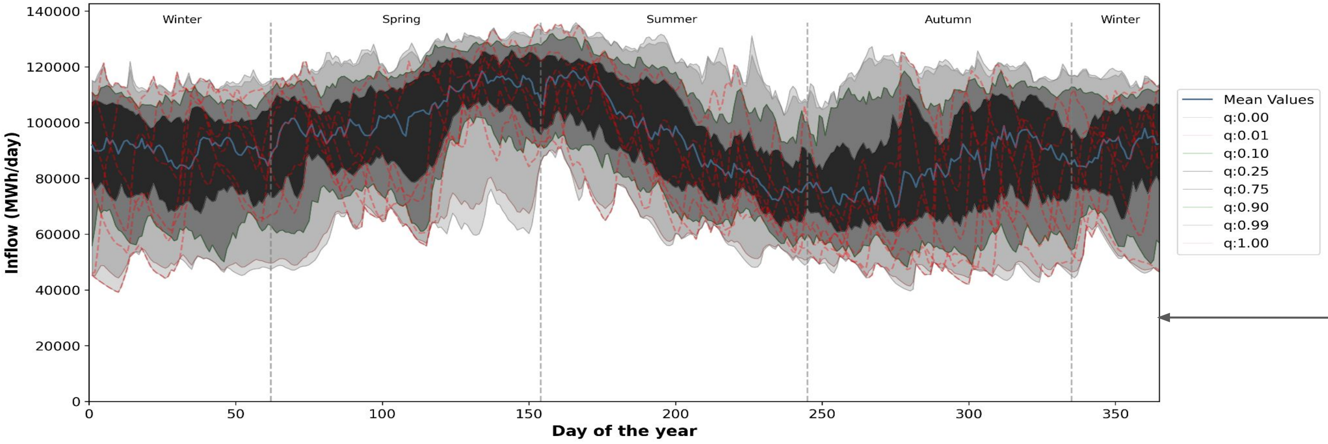
A **K-Fold cross-validation** was chosen despite the time-series nature of the problem since we are interested in **back-casting and re-projecting**, and not strictly forecasting.

Distributional shift errors due to generalization to **CORDEX** data when projecting are not measurable since there is no ground truth available, thus care should be taken when analyzing the projection results.



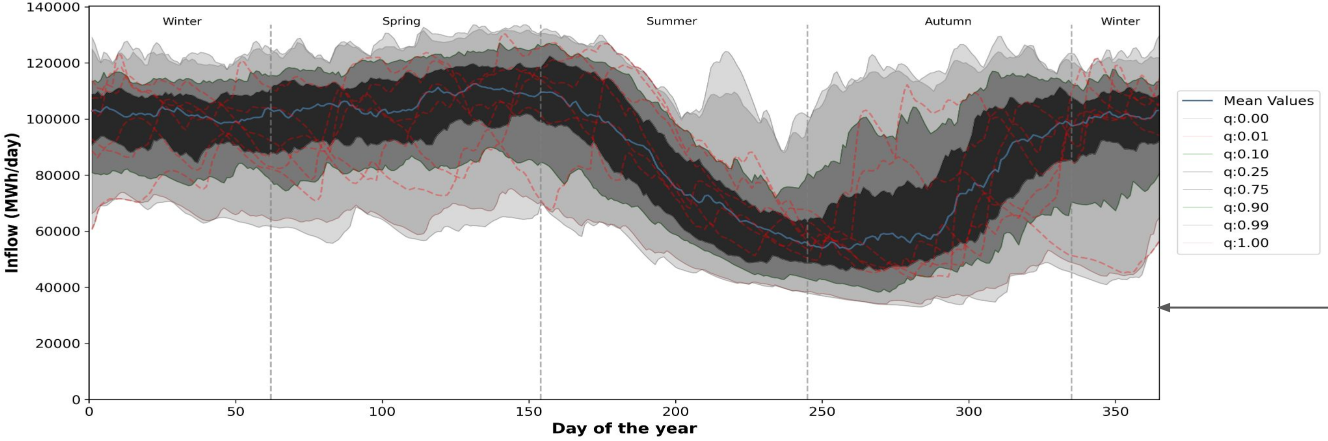


HYDRO INFLOW RESULTS



The model **generalizes** the seasonal dynamics, allowing to estimate inflow for the **EFAS** data period of **1991-2021**, starting from data in 2010-2021.

Historical (EFAS)



Generalizing with **CORDEX** (**2006-2065**) often results slower dynamics and changes in seasonality. Care should be taken quantifying the impact of **climate change** vs model **distributional shift**.

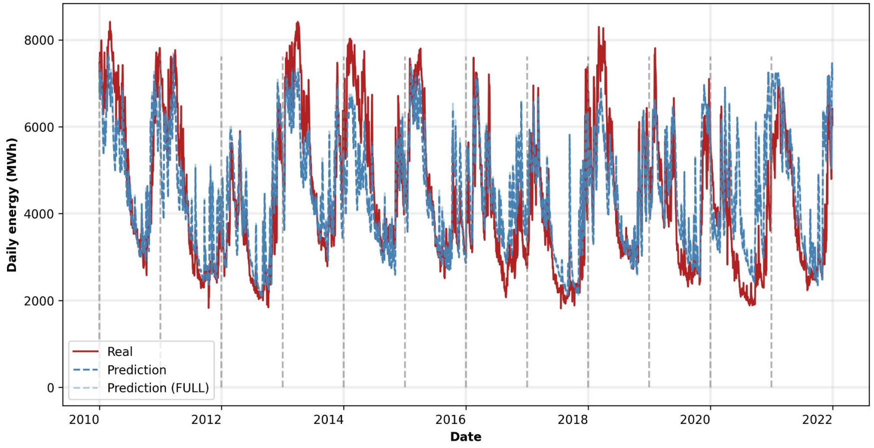
Projections (CORDEX)

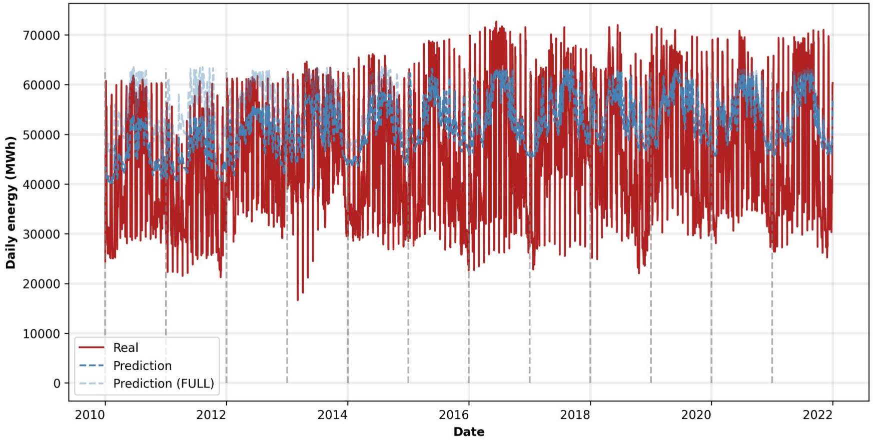
AUTOMATED TYPE IDENTIFICATION

The model is **robust** when **Run of River PP show "regulated" weekly dynamics**, exhibiting consistency with natural inflows, avoiding to fit the non-natural components of the generation signal.

Still, a data-driven classification method may prove beneficial in preventing such occurrences.

Moreover, **mis-classification of the PP behaviour** can impact studies that need to quantify zone-wide dispatching capabilities and generation correlation with demand, as **adequacy studies**.





AUTOMATED TYPE IDENTIFICATION

Hydroelectric power plants are broadly categorized into 4 types:

- **Run of River** (no storage capabilities)
- **Pondage** (up to 24h of storage)
- **Reservoir** (more than 24h of storage)
- **Open loop pumping** (reservoir with pumping)
- **Closed loop pumping** (pumping with no natural inflow)

We propose an alternative surrogate **data-driven classification** that could be used:

- When handling **complex datasets** from various sources and the labeling could me missing
- To check for **mislabeling**
- For **PP in sequence**, e.g. a RoR after a Reservoir, behaving as a Reservoir.

AUTOMATED TYPE IDENTIFICATION

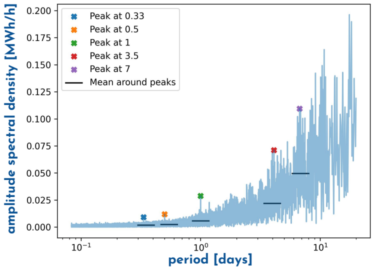
We propose the following approach:

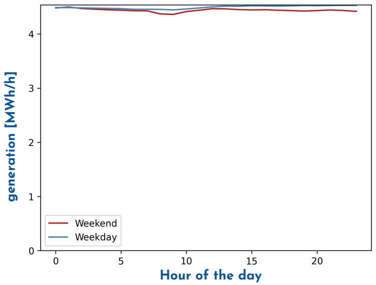
- Compute the **Amplitude Spectral Density** of the hourly generation signal
- Evaluate the **peaks prominences** at some key periods:
  - **1, 1/2, 1/3 days** → periods associated with daily regulation
  - **7, 3.5 days** → periods associated with weekly regulation

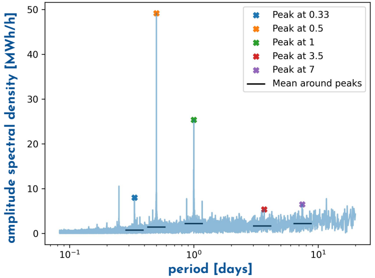
AUTOMATED TYPE IDENTIFICATION

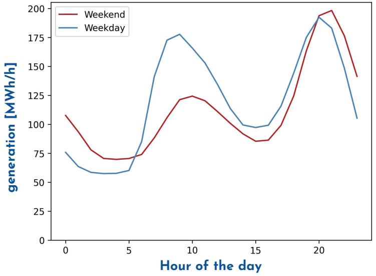
These are few examples of **Amplitude Spectral Density** of power plants with different behaviours.

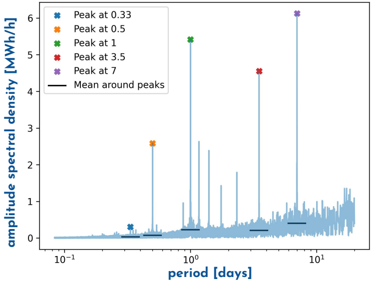
While similar observations could be obtained by looking at the **hour-by-hour-by-weekday** aggregation, the ASD provides **useful scalar values** that allows to systematically analyze large amount of generation data.

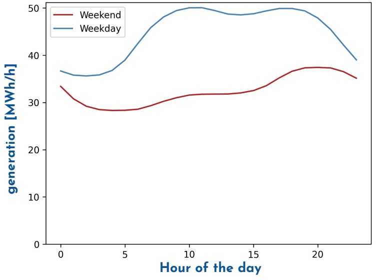






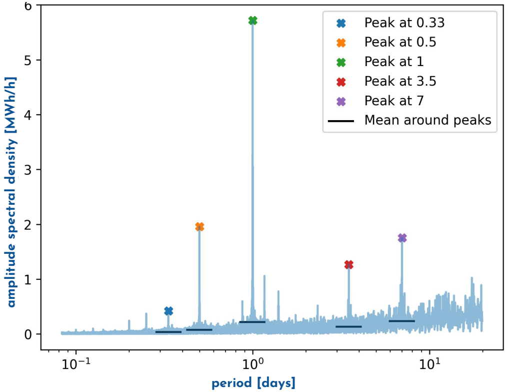




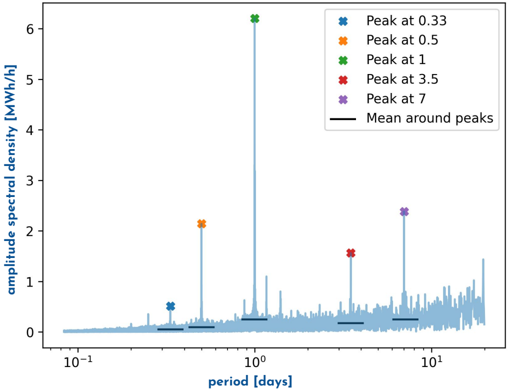


AUTOMATED TYPE IDENTIFICATION

Upstream reservoir and downstream RoR interferences



Upstream Reservoir PP:
This ASD is due to the PP behaviour, with 6 MW harmonics on the 1 day period, and 2 MW on the 1 week one.



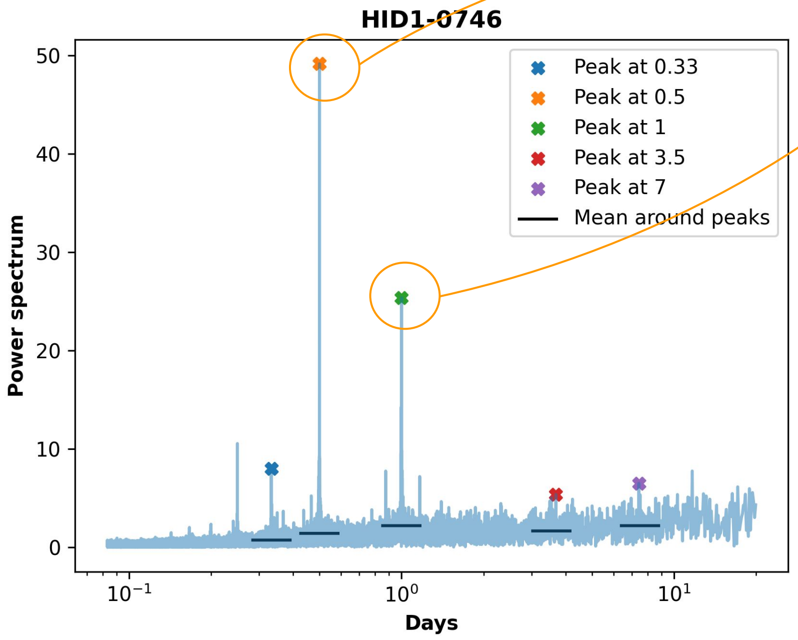
Downstream RoR PP:
This ASD is due to the upstream reservoir behaviour, since the PP has no modulation capabilities.

The end.

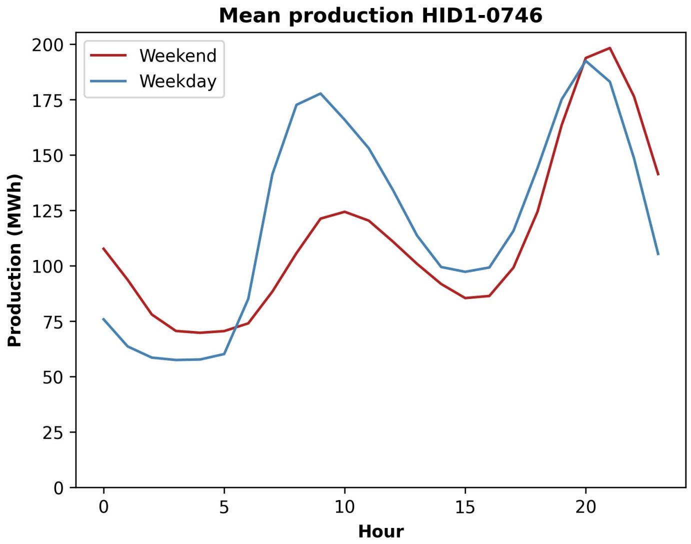
Thanks for listening.

Back-up

FOURIER ANALYSIS - POWER SPECTRUM



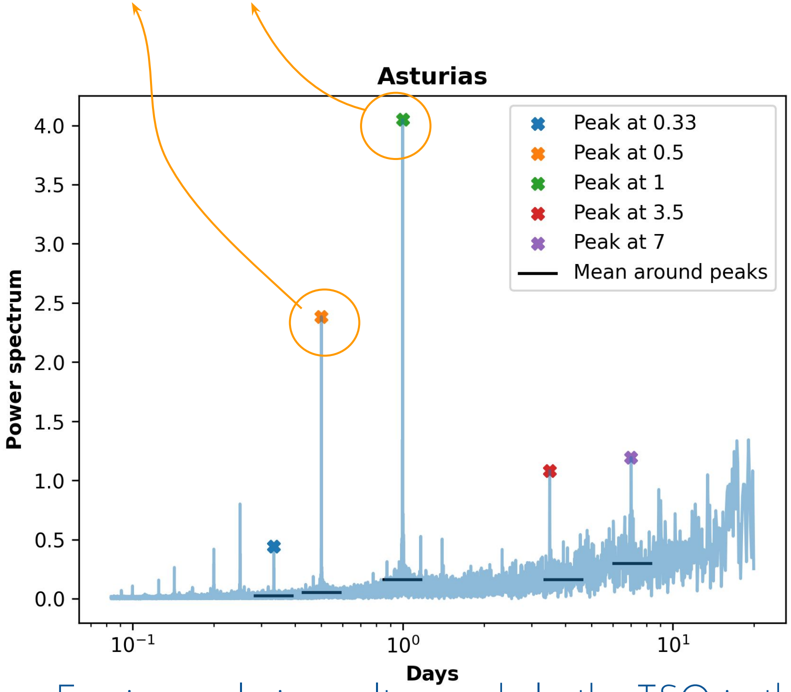
Reservoir daily and intra daily modulation

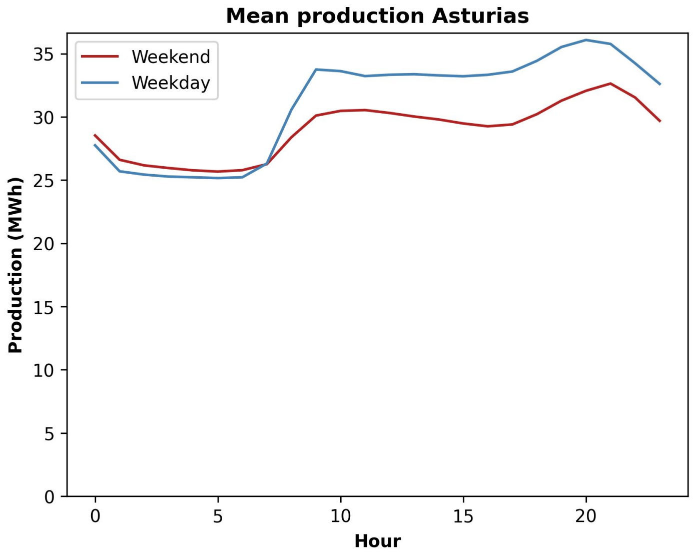


FOURIER ANALYSIS - POWER SPECTRUM

RoR with modulation

Asturias PP is classified as RoR, but there is some modulation. Are there upstream modulations? Kept as RoR or not?





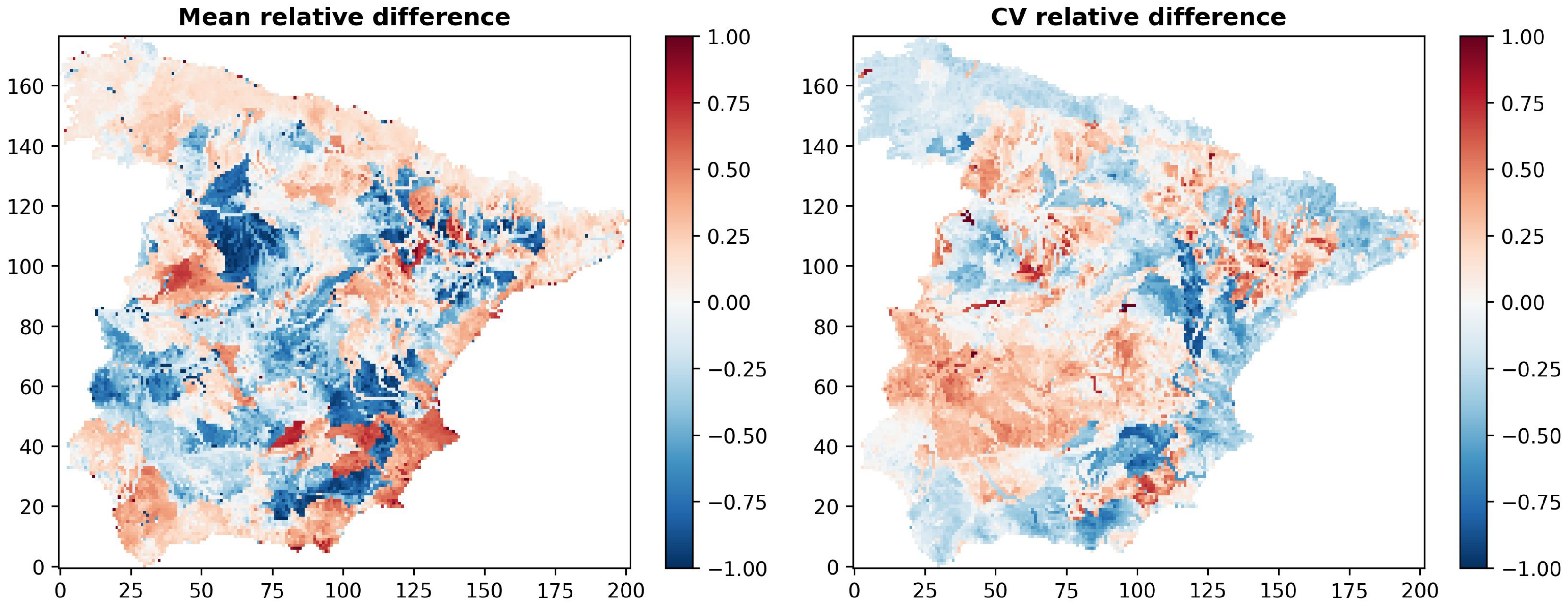
Fourier analysis results may help the TSO in the definition of the PPs' type (e.g. RoR, Pondage, Reservoir).

EFAS (vs) CORDEX

Maps of the difference EFAS - CORDEX in Spain 2006-2021

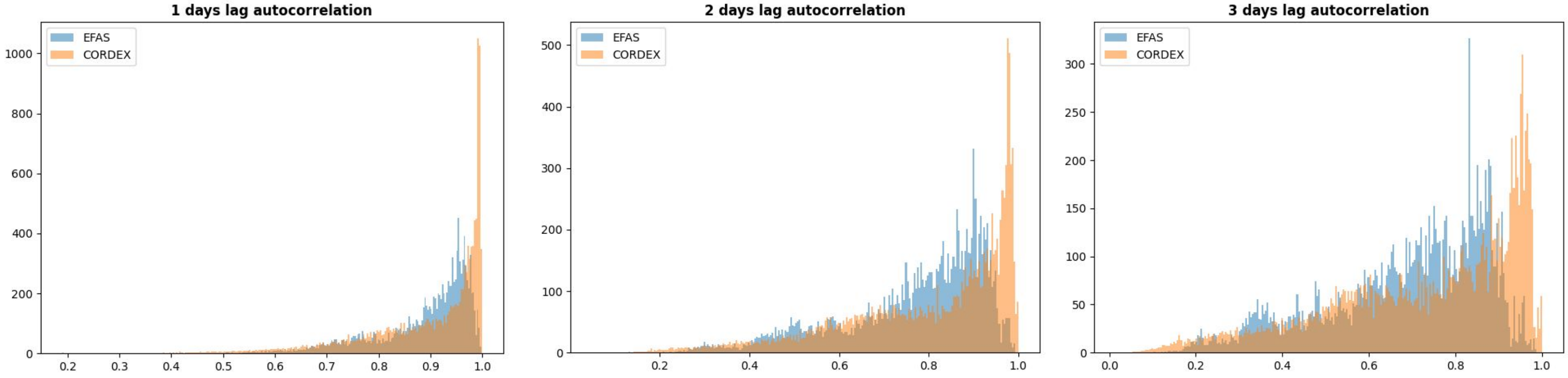
$$\frac{(efas_{mean} - cordex_{mean})}{(efas_{mean} + cordex_{mean})}$$

$$\frac{\left(\frac{efas_{std}}{efas_{mean}} - \frac{cordex_{std}}{cordex_{mean}}\right)}{\left(\frac{efas_{std}}{efas_{mean}} + \frac{cordex_{std}}{cordex_{mean}}\right)}$$



EFAS (vs) CORDEX

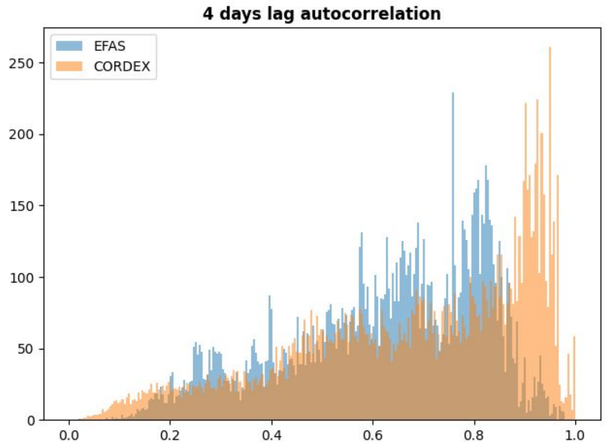
Autocorrelations of all the values of Spain, weighted by their logarithmic mean 2006-2021

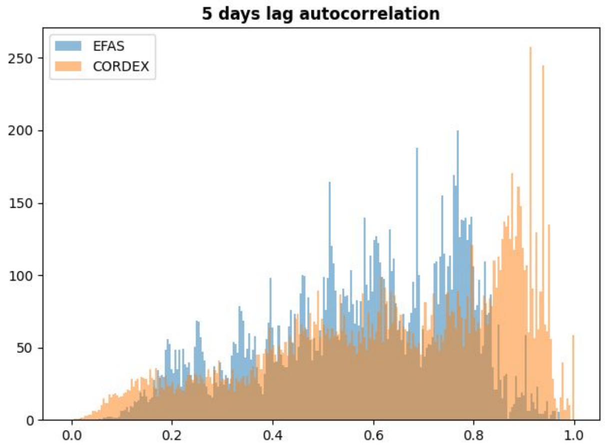


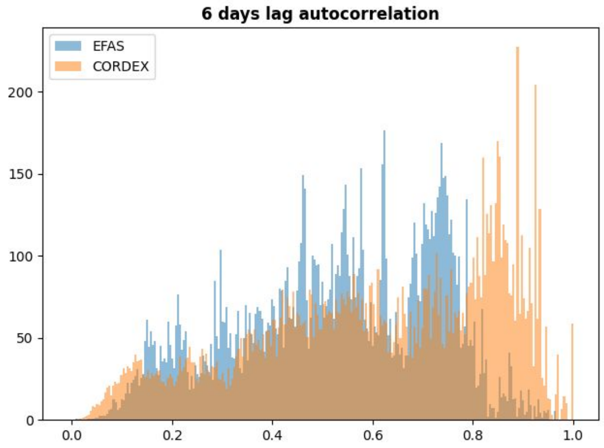
→ Larger fractions of extremely autocorrelated pixels
→ Different behaviour in frequency between EFAS and CORDEX

EFAS (vs) CORDEX

Autocorrelations of all the values of Spain, weighted by their logarithmic mean 2006-2021





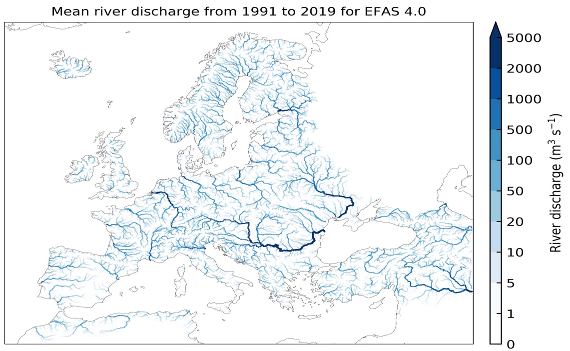


→ Larger fractions of extremely autocorrelated pixels
→ Different behaviour in frequency between EFAS and CORDEX

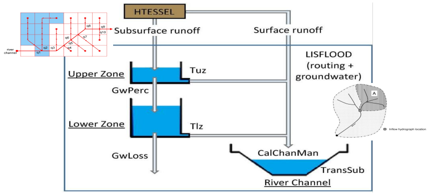
HISTORICAL RIVER DISCHARGE (EFAS)

Daily and **6-hourly** discharge time series **for every grid** cell of the river network.

- 5 km grid
- From 1991 to today
- uses **LISFLOOD** hydrological model
- download by **browser** or **Python** API
- **gridded** data
- delivering as **.grib** and **.netcdf**
- **free of charge** both **historic** and **forecast**



LISFLOOD is a Rainfall-runoff model capable of simulating the hydrological processes that occur in a catchment.



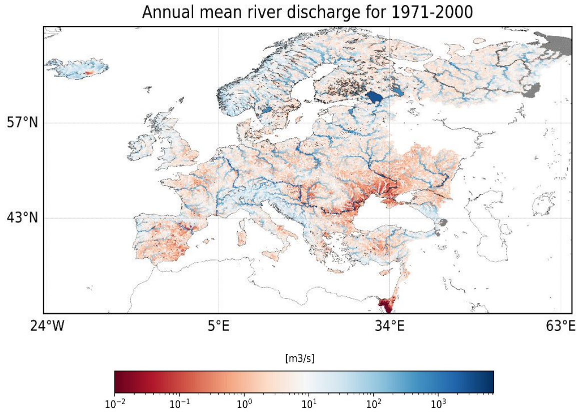
https://ec-jrc.github.io/lisflood-model/1_1_introduction_LISFLOOD/

CORDEX

Daily discharge time series **for every grid** cell of the river network.

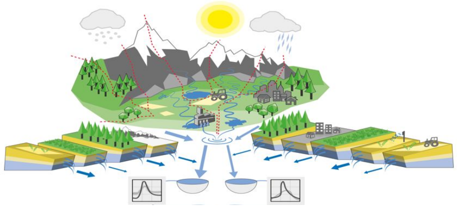
- 5 km grid
- From 2006 (or 1971*) to 2100
- uses **E-HYPEgrid** hydrological model
- download by **browser** or **Python** API
- **gridded** data
- delivering as **.grib** and **.netcdf**
- **free of charge** both **historic** and **forecast**

**Depending on the model.*



The Hydrological Predictions for the Environment (HYPE) is a physically based catchment model, which simulates water flow and substances on their way from precipitation through different storage compartments and fluxes to the sea.

https://climate.copernicus.eu/user-guidance



CORDEX

River discharge CORDEX data are produced according to the Representative Concentration Pathway (**RCP**) **4.5** scenario.

In this scenario, the employment of technologies and strategies for reducing greenhouse gas emissions would allow to stabilise the radiating forcing at 4.5 W/m^2 before the year 2100.