From data mining to information extraction: using ERA-INTERIM reanalysis to model hydro-power production in Europe



Climate Change

Matteo De Felice ⁽¹⁾, L. Dubus ⁽²⁾, A. Troccoli ⁽³⁾ (1) ENEA, Italy (2) EDF R&D, France (3) Univ. of East Anglia, UK







why?



- Hydro-power is still the dominant RES in many countries: in EU-28 in 2015 the largest RES accounting for 14.4% of total primary energy production [source: EUROSTAT]
- It is a flexible source: it's crucial in high RES penetration scenarios





Installed capacity in Europe





We have three types of HP plants:

- 1. Run-of-river
- 2. Storage dams
- 3. Pumped storage

Data: ENTSO-E Power Statistics





Challenges

- Modelling river flow from weather is hard: need to model precipitation, run-off, evaporation, snow melt, etc.
- Water in reservoirs can be used for multiple purposes, e.g. irrigation.
- There are other constraints, such as maintaining navigability on rivers, fish ladders, water levels for recreation, water cooling for thermal power plants, etc.
- We need information at basin-level!





Data source(s)

- ENTSO-E data on hydro-power generation available on the **Transparency Platform.**
- Two types: 1) Installed capacity (Installed Generation Capacity) ulletAggregated, 14.1.A) and 2) Hourly generation data (Aggregated Generation per Type, 16.1.B&C).
- Three typologies: pumped storage, run-of-river and poundage, \bullet and water reservoir.
- Data available since 1/1/2015 and for this work we have used ullet24 months of data (until 31/12/2016).
- Meteorological data (precipitation, temperature and snow depth) ullethas been extracted from the ERA-INTERIM





predictors h e

- **Country-level averages** ۲
- Daily data ullet
- Temperature, precipitation and • snow-depth (ERA-INTERIM)







lagged predictors

Main question: is the connection between meteorological predictors and generation instantaneous?

- Rolling sum of the last N days to predict generation at day t
- What is the best value of N?
- Our approach: find the lag that maximise the correlation between the rolling sum and the generation



Optimising the lagged predictors

Climate Change

Correlation between generation and cumulative sum of variables

FR, country production









Lagged predictors

Correlation between generation and cumulative sum of variables



IT, country production

ENTSO-E Data 01/01/2015-30/06/2017, ERA-INTERIM Data



opes eyes on caro





Results



Out-of-sample correlation





Results – France, run-of-river

0.35 0.30 0.25 0.20 0.15 0.10 2015-01 2015-07 2016-01 2016-07 2016-07 2016-07 2016-07 2016-07 2017-01

variable — rf_out — target





Results – France, run-of-river (nolag)









Results-Spain, reservoir



variable — rf_out — target





Results-Spain, reservoir (nolag)



variable — rf_out — target





Takeaway messages

- Data-driven and generalised modelling of HP generation: (surprisingly?) good results
- (Another) building block of a simulator of the European power networks
- Climate-focused analysis on the historical data
- Possibility to use the model to perform predictions?

