Characterizing the Spanish hydro basins for their use in openTEPES

Jesús David Gómez Pérez

Andrés Ramos Galán

World openTEPES Conference

2nd World openTEPES Conference | Madrid, 19th of July 2023

openTEPES

IIT INSTITUTO DE INVESTIGACIÓN TECNOLÓGICA

COMILLAS ICAI

1. Introduction
   - The Significance of Hydropower Generation
   - Challenges in Modeling Hydropower Generation

2. Methodology
   - Our first approximation: Regression models
   - Model Training and Validation

3. Results
   - Results
   - Model Interpretation and Insights

4. Conclusions
   - Future directions and challenges

World openTEPES Conference

1. **Introduction**

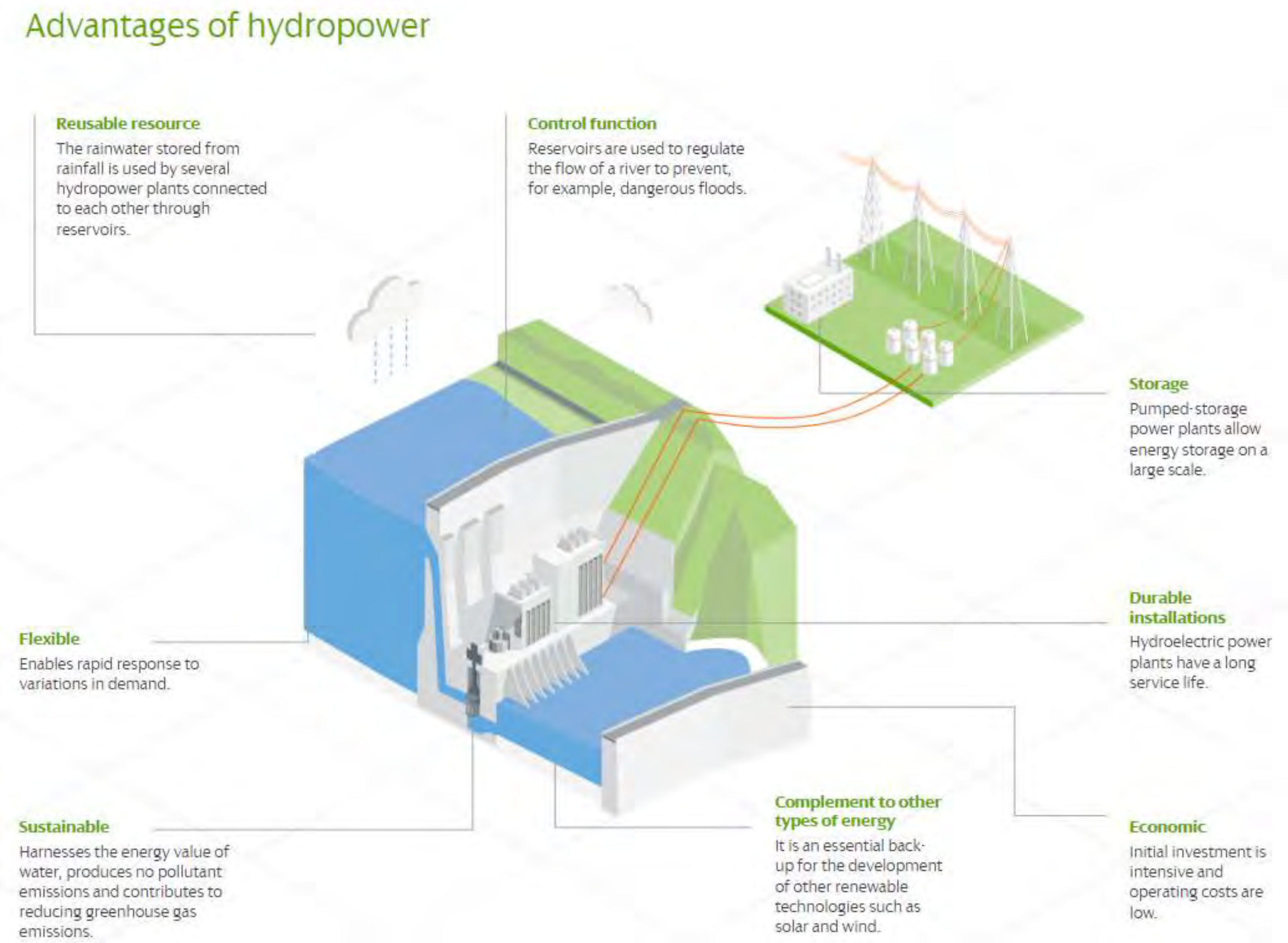
2. Methodology

3. Results

4. Conclusions

World openTEPES Conference

The Significance of Hydropower Generation



Source: Iberdrola. https://www.iberdrola.com/about-us/what-we-do/hydroelectric-power

Challenges in Modeling Hydropower Generation

REAL LIFE	MATHEMATICAL MODEL
ENVIRONMENTAL CONSTRAINTS	• Natural water inflows uncertainty
	• Minimum and maximum power and reservoir levels
OPERATIONAL CONSTRAINTS	• Minimum number of units on line
	• Upward and downward ramping rates
	• Reservoir level restrictions
REGULATORY CONSTRAINTS	• Total release volume

And if we want to conduct a planning study, considering the model configured in openTEPES, how can we incorporate these features?

1. Introduction

2. Methodology

3. Results

4. Conclusions

World openTEPES Conference

Our first approximation: Regression models

Modeling dynamic minimum and maximum power and dynamic upward and downward ramping rates as a function of other variables

What do we have?

- ESIOS
  - Programming units
  - UGH* production. P48Cierre
  - Wind and solar production
  - Demand
- REE
  - Hydro producible
- MITECO
  - State of basins

What do we expect?

$$P_{max} = f(\ldots)$$

$$P_{min} = f(\ldots)$$

$$\frac{dP}{dt} = f(\ldots)$$

→ openTEPES

- Endogenous model → Linear model
- Exogenous model → Linear or more complex model

* UGH: “Unidad de gestión hidráulica” (Hydro management unit)

Model training and validation

Predictor variables

NUMERICAL

- Producible
- Wind / Demand
- Solar / Demand
- Reserve Level
- Lags of production

CATEGORICAL

- Number of week

Data set
Production: 2011 – 2019 hourly
Producible: 1991 – 2023 monthly
Wind, solar, and demand: 2012 – 2020 monthly
Water reserve level: 1988 – 2023 weekly
Training set: 70%
Test set: 30%

Linear model	Linear or more complex model
• OLS • Bayesian linear regression (OLS with Lasso)	• Neural networks (MLP)

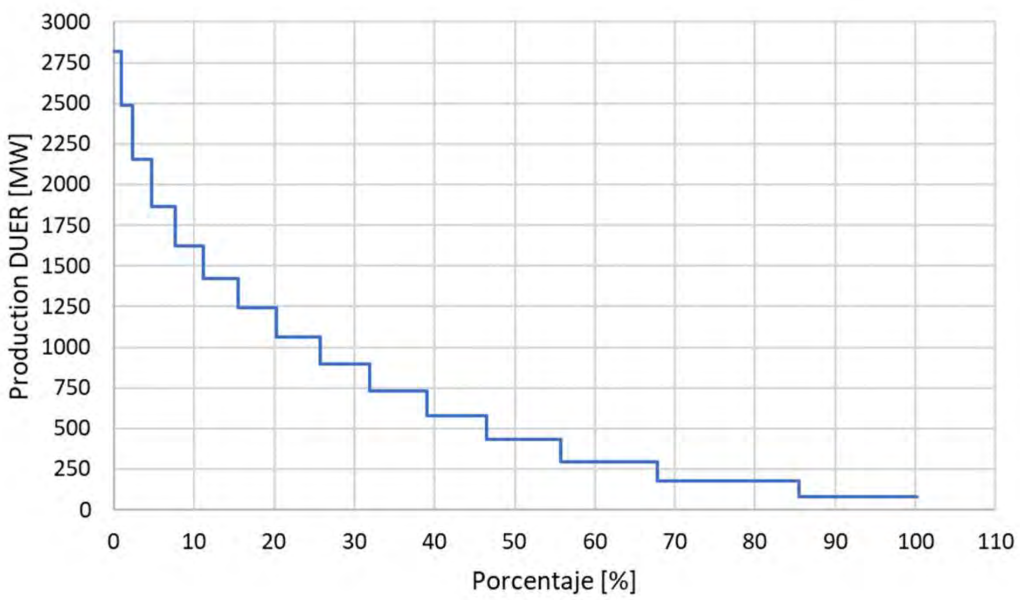
* UGH: "Unidad de gestión hidráulica" (Hydro management unit)

“

World openTEPES Conference

Minimum and maximum power

What production value should we choose to model the minimum and maximum in an operational context?



Procedure (duration curve):

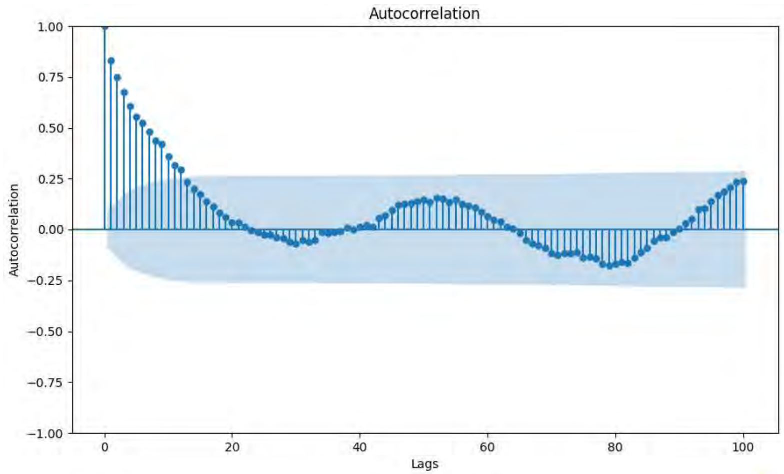
1. The production will be clustered into 15 groups
2. Plot the empirical probability distribution for each centroid
3. Select the first and last clusters as Pmax and Pmin, respectively

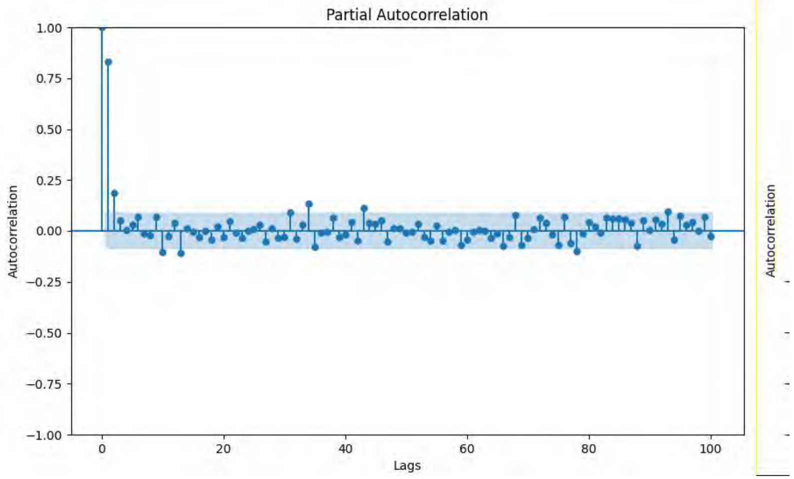
DUER:
Pmax = quantil 99
Pmin = quantil 15

Linear regression analysis

	DUER	EBRFEN	GDLQ	GDNA	JUCA	SIL	TAJO	B_CANTABRICO	B_DUERO	B_EBRO	B_GALICIACOSTA	B_GUADALQUIVIR	B_GUADIANA	B_JUCAR	B_SIL	B_TAJO	relative_reserve	relative_producible	solar_demand_portion	wind_demand_portion
DUER	1.00	0.39	0.18	-0.06	0.10	0.62	0.45	0.17	0.45	0.34	0.33	0.46	0.44	0.30	0.40	0.42	0.53	0.59	-0.29	0.36
EBRFEN	0.39	1.00	0.51	0.26	0.22	0.34	0.40	0.65	0.63	0.79	0.66	0.55	0.50	0.49	0.70	0.69	0.68	0.57	0.23	0.20
GDLQ	0.18	0.51	1.00	0.67	0.47	0.04	0.43	0.50	0.51	0.53	0.42	0.54	0.52	0.50	0.56	0.61	0.56	0.17	0.46	-0.17
GDNA	-0.06	0.26	0.67	1.00	0.42	-0.17	0.15	0.36	0.38	0.33	0.21	0.36	0.36	0.44	0.37	0.39	0.33	-0.20	0.53	-0.38
JUCA	0.10	0.22	0.47	0.42	1.00	0.01	0.23	0.12	0.31	0.28	0.27	0.36	0.28	0.34	0.29	0.27	0.34	-0.07	0.33	-0.19
SIL	0.62	0.34	0.04	-0.17	0.01	1.00	0.35	0.19	0.24	0.22	0.33	0.26	0.20	0.19	0.30	0.23	0.37	0.65	-0.41	0.56
TAJO	0.45	0.40	0.43	0.15	0.23	0.35	1.00	0.17	0.42	0.40	0.29	0.45	0.45	0.28	0.33	0.55	0.49	0.36	-0.07	0.11
B_CANTABRICO	0.17	0.65	0.50	0.36	0.12	0.19	0.17	1.00	0.49	0.73	0.79	0.17	0.11	0.19	0.80	0.49	0.42	0.42	0.48	0.13
B_DUERO	0.45	0.63	0.51	0.38	0.31	0.24	0.42	0.49	1.00	0.79	0.64	0.69	0.69	0.57	0.79	0.91	0.93	0.32	0.40	-0.03
B_EBRO	0.34	0.79	0.53	0.33	0.28	0.22	0.40	0.73	0.79	1.00	0.78	0.51	0.46	0.41	0.78	0.76	0.73	0.42	0.39	0.10
B_GALICIACOSTA	0.33	0.66	0.42	0.21	0.27	0.33	0.29	0.79	0.64	0.78	1.00	0.27	0.20	0.17	0.91	0.51	0.60	0.55	0.35	0.23
B_GUADALQUIVIR	0.46	0.55	0.54	0.36	0.36	0.26	0.45	0.17	0.69	0.51	0.27	1.00	0.97	0.88	0.41	0.74	0.84	0.27	0.09	0.03
B_GUADIANA	0.44	0.50	0.52	0.36	0.28	0.20	0.45	0.11	0.69	0.46	0.20	0.97	1.00	0.86	0.37	0.76	0.83	0.24	0.08	0.01
B_JUCAR	0.30	0.49	0.50	0.44	0.34	0.19	0.28	0.19	0.57	0.41	0.17	0.88	0.86	1.00	0.32	0.62	0.73	0.18	0.20	0.06
B_SIL	0.40	0.70	0.56	0.37	0.29	0.30	0.33	0.80	0.79	0.78	0.91	0.41	0.37	0.32	1.00	0.69	0.73	0.52	0.46	0.12
B_TAJO	0.42	0.69	0.61	0.39	0.27	0.23	0.55	0.49	0.91	0.76	0.51	0.74	0.76	0.62	0.69	1.00	0.90	0.35	0.38	-0.01
relative_reserve	0.53	0.68	0.56	0.33	0.34	0.37	0.49	0.42	0.93	0.73	0.60	0.84	0.83	0.73	0.73	0.90	1.00	0.49	0.24	0.14
relative_producible	0.59	0.57	0.17	-0.20	-0.07	0.65	0.36	0.42	0.32	0.42	0.55	0.27	0.24	0.18	0.52	0.35	0.49	1.00	-0.34	0.71
solar_demand_portion	-0.29	0.23	0.46	0.53	0.33	-0.41	-0.07	0.48	0.40	0.39	0.35	0.09	0.08	0.20	0.46	0.38	0.24	-0.34	1.00	-0.52
wind_demand_portion	0.36	0.20	-0.17	-0.38	-0.19	0.56	0.11	0.13	-0.03	0.10	0.23	0.03	0.01	0.06	0.12	-0.01	0.14	0.71	-0.52	1.00

Exogenous model





Linear model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DUER   R-squared:                       0.730
Model:                            OLS   Adj. R-squared:                  0.726
Method:                 Least Squares   F-statistic:                     173.4
Date:                Mon, 17 Jul 2023   Prob (F-statistic):           6.09e-89
Time:                        19:32:11   Log-Likelihood:                -248.89
No. Observations:                 327   AIC:                             509.8
Df Residuals:                     321   BIC:                             532.5
Df Model:                           5
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                    0.0058      0.029      0.201      0.841      -0.051       0.063
relative_producible      0.2339      0.050      4.714      0.000       0.136       0.331
wind_demand_portion     -0.0751      0.044     -1.690      0.092      -0.162       0.012
solar_demand_portion    -0.0990      0.034     -2.954      0.003      -0.165      -0.033
DUERlag1                 0.5742      0.056     10.258      0.000       0.464       0.684
DUERlag2                 0.1491      0.054      2.737      0.007       0.042       0.256
==============================================================================
Omnibus:                        7.515   Durbin-Watson:                   1.946
Prob(Omnibus):                  0.023   Jarque-Bera (JB):               12.267
Skew:                           0.031   Prob(JB):                      0.00217
Kurtosis:                       3.947   Cond. No.                         4.32
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
R^2 score linear regression in the test set:  0.7525112471393269
```

- R^2 Training set: 0.73
- R^2 Test set: 0.75

Significant variables:

- Producible
- Solar/demand
- Lag 1
- Lag 2

Wind/demand (p-value > 0.05)

Bayesian linear model (Lasso)

```
R^2 Lasso:  0.7091400496239069
R^2 Lasso in the test set:  0.7417066642015798
coefficients lasso:  [ 0.14425389  0.          -0.00234461  0.55955267  0.10291472]
intercept lasso:  0.008730207738090592
```

- R^2 Training set: 0.709
- R^2 Test set: 0.742

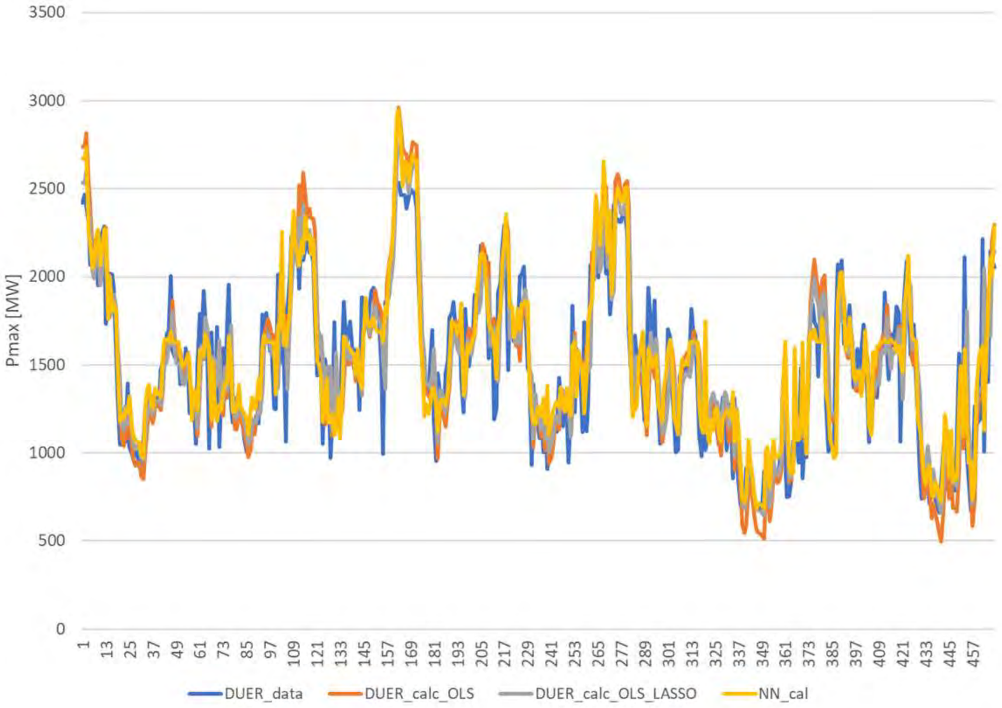
Neural Networks (MLP)

```
R^2 score neural network:  0.8451768864529361
5/5 [==============================] - 0s 750us/step
R^2 score neural network in the test set:  0.6718192231708278
```

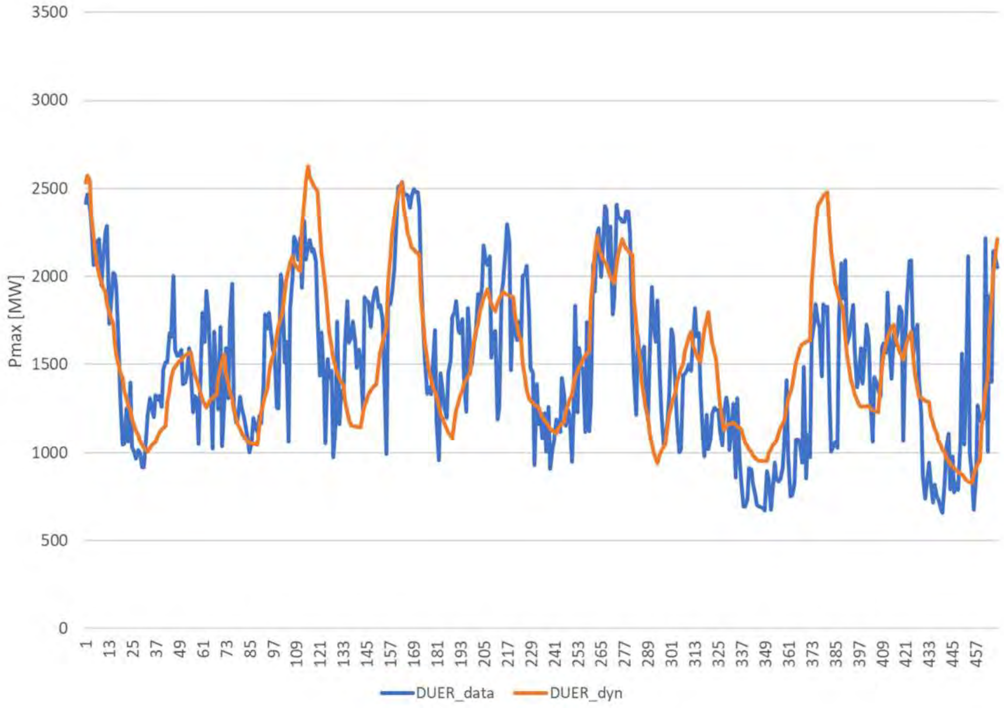
- R^2 Training set: 0.845
- R^2 Test set: 0.672

Model interpretation and insights

Different models



Lasso (Dynamic)



World openTEPES Conference

Conclusions

- Modeling the time-dependent maximum and minimum power, as well as the ramp-up and ramp-down rates, is a first approximation to capture many of the constraints of hydropower
- The duration curve is a useful tool to determine the percentile associated with the minimum and maximum values of a variable
- Linear models adequately represent the objective of this study
- *How to improve the prediction of models that considers temporal correlation?*

Thank you

Jesús David Gómez Pérez

jdgomez@comillas.edu

Break 10 min

World openTEPES Conference