Modeling of a Fully Reneweable Energy grid with Hydrogen Storage: a stochastic approach considering time interdependence of Wind and Solar Power.

Gabor Riccardi[1], Bianca Urso[2],
Advisor: Stefano Gualandi
[1] University of Pavia, [2] IUSS - School for Advanced Studies, Pavia

Abstract

In recent years, the integration of renewable energy sources into electrical grids has become a critical area of research due to the increasing need for sustainable and resilient energy systems. In this report we present a comprehensive model for an electrical grid powered by wind and photovoltaic (PV) systems, supported by hydrogen storage. The main contributions of this study are the following: In the first section we deal with the scenario generation (SG) process for wind and PV power output, acknowledging the inherent dependencies between these variables over time. These dependencies are captured using marginal distributions coupled with a Gaussian copula, ensuring that the generated scenarios realistically reflect the temporal correlations observed in historical data.

In the second section, we model the structure of a hypothetical electrical grid as a capacity expansion model. To ensure the problem is tractable, we solve it by iteratively refining the temporal aggregation. Furthermore, we provide a theoretical justification for why this method can be used to efficiently warm start each iteration. The third section we show the results of various numerical experiments.

Introduction

The threat of climate change is pushing policy-makers to pursue greater integration of renewable energy sources into electrical grids, while at the same time ensuring reliability and resilience of the grids. The model presented in this report explores the possibility of an electrical grid powered entirely through wind and photovoltaic (PV) systems, and supported by hydrogen storage. It is of interest to estimate the power generation capacity for both wind and solar, as well as the hydrogen storage and conversion capacities, that would be necessary in order to power a reliable grid supplying residential electricity load and industrial base load of both electricity and hydrogen for a given area, while minimizing the cost of implementing such infrastructure.

The main difficulties arising when designing such a system lie in the great variability of the generation of electricity through wind and solar, since these resources are highly dependent on weather conditions, making it impossible to plan long-term by optimizing on forecasts, and requiring a statistical approach to generate realistic scenarios. In [PK09], Papaefthymiou et al. model the interdependence between stochastic variables in renewable generation with a Gaussian copula. We extend this approach to generate realistic scenarios for wind and solar power generation for long time horizons. The construction of the PDF for wind and solar is explained in detail in subsections 1.1.1 and 1.1.2 respectively. While fitting on historical data we did not account for possible changes

in future climate. The grid is modeled as a two stage stocastic problem as described in section 2. In the first stage, the capacity expansion of each generator, hydrogen storage, transmission line, and hydrogen pipe is decided. In the second stage, the economic dispatch is solved for each scenario. Computational costs increase rapidly with the number of nodes and scenarios. To adress this, in [PJ24] Pecci et al. propose a regularized decomposition method. In [Hör+18], Brown et Al use a node aggregation method that allows to reduce the computational costs. In a similar direction we introduce an iterative algorithm that starts with an initial time aggregation. For example, instead of considering hourly demand and production, we aggregate the scenarios into weekly demand and production. This initial aggregation corresponds to relaxing the original model. Next, we identify which time steps have violated constraints and split the corresponding intervals into two, thereby obtaining a finer time partition. The corresponding CEP model is then redefined, tightening the relaxation. The previous solution is used to warm start the new iteration. The algorithm is described in detail in section 3

1 Scenario Generation

Wind and solar power are inherently intermittent and uncertain, posing a challenge to their successful integration into the energy system. Hydrogen storage and other forms of energy storage offer potential solutions to mitigate these issues. However, the amount of long-term storage required in a fully renewable grid is heavily influenced by the stochastic behavior of wind and solar power. Moreover, historical data typically covers only a limited number of climate years, which restricts the ability to test the grid over long time horizons encompassing various climatic conditions. To address this limitation, we adopted a scenario generation (SG) method based on historical data of each of the European countries considered, allowing us to create realistic and diverse scenarios that better capture the variability and uncertainty of renewable energy sources over extended periods.

To model the probability distribution corresponding to the power output of wind turbines for each hour of the year, we utilized a Weibull distribution, justified by its proven effectiveness in capturing the variability and skewness of wind power distributions [[Kha+14]]. For solar power, a Beta distribution was employed in [[Yua+19]]. To account for interdependence between temporally near time steps, we coupled these distributions using a Gaussian Copula approach, which captures the dependencies between hourly power outputs effectively. This approach accurately mimics common weather phenomena.

1.1 Stochastic Processes description

The stochastic processes of power observations will be denoted as Y_t. Where $t \in T$, is the set indexing all the random variables which want to be considered jointly. We assume that the random variable Y_t has either a Weibull distribution, in the case of Wind Power, or a Beta distribution in the case of Solar Power. The electricity load is taken from the [24]. We normalised the 2023 data by country to indicate the trend of load throughout the year, dividing by mean hourly load: this is then to be multiplied by the mean load of the area the policy maker is interested in serving with the modeled grid.

1.1.1 Parametric Estimation of Wind Power distribution

The parameters defining the Weibull Distribution are estimated using the Maximum Likelyhood Estimation. The Weibull density function is given by:

$$f(x;\theta,\gamma) = \left(\frac{\gamma}{\theta}\right) x^{\gamma-1} \exp\left(-\left(\frac{x}{\theta}\right)^{\gamma}\right)$$

where $\theta, \gamma > 0$ are the scale and shape parameters, respectively. Given observations $X_1, \ldots, X_n$, the log-likelihood function is:

$$\log L(\theta,\gamma) = \sum_{i=1}^{n} \log f(X_i \mid \theta,\gamma)$$

The optimum solution is found by searching for the parameters for which the gradient is zero :

$$\frac{\partial \log L}{\partial \theta} = -\frac{n\gamma}{\theta} + \frac{\gamma}{\theta^2} \sum_{i=1}^{n} x_i^{\gamma} = 0 \tag{1}$$

Eliminating θ, we get:

$$\left[\frac{\sum_{i=1}^{n} x_i^{\gamma} \log x_i}{\sum_{i=1}^{n} x_i^{\gamma}} - \frac{1}{\gamma}\right] = \frac{1}{n} \sum_{i=1}^{n} \log x_i \tag{2}$$

This can be solved to get the MLE estimate $\hat{\gamma}$. This can be accomplished with the aid of standard iterative procedures such as the Newton-Raphson method or other numerical procedures. This is done with the aid of the package *scipy*. Once $\hat{\gamma}$ is found, $\hat{\theta}$ can be determined in terms of $\hat{\gamma}$ as:

$$\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\hat{\gamma}}\right)^{\frac{1}{\hat{\gamma}}} \tag{3}$$

1.1.2 Parametric Estimation of Solar Power distribution

To estimate the α and β parameters defining the Beta distribution Y, we use the *Method of Moments*. The mean of the random variable Y can be expressed as $\mathbb{E}[Y] = \frac{\alpha}{\alpha+\beta}$ and the variance as $Var[Y] = \frac{\alpha+\beta}{(\alpha+\beta)(\alpha+\beta+1)}$. In particular by explicating β in the first equation and substituting it in the second equation we obtain that:

$$\begin{cases} \alpha = \mathbb{E}[X]\left(\frac{\mathbb{E}[X](1-\mathbb{E}[X])}{\mathrm{Var}[X]} - 1\right) \\ \beta = (1-\mathbb{E}[X])\left(\frac{\mathbb{E}[X](1-\mathbb{E}[X])}{\mathrm{Var}[X]} - 1\right) \end{cases} \tag{4}$$

By substituting the mean and the variance with their empirical approximation we obtain the method of moments estimator for α and β.

1.1.3 Parametric Copula Estimation

The cumulative density function of both the Weibull and Beta distributions are continuous and invertible. Therefore, the random variables $U_t \coloneqq F_{Y_t}(Y_t)$ have a uniform distribution over $[0,1]$. The copula of the random variables $\{Y_t\}_{t\in T}$ is defined as the function $C : [0,1]^T \to [0,1]$ such that

$$C(F_{Y_1}(y_1), \dots, F_{Y_T}(y_{|T|})) = P(Y_1 \leq y_1, \dots, Y_{|T|} \leq y_{|T|}). \tag{5}$$

This function always exists because of Sklar's Theorem. The Gaussian Copula represents well the coupled behavior in renewable stochastic systems [[PK09]] and is the one used in this project. For a given correlation matrix Σ, the Gaussian Copula with parameter matrix Σ is defined as $C_\Sigma^{\text{Gauss}}(u_1, \dots, u_T) \coloneqq \Phi_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_T))$. Where Φ, Φ_Σ are the cdf Gaussian variables having distribution $\mathcal{N}(0,1)$ and $\mathcal{N}(0,\Sigma)$ respectively. In particular if C_Σ^{Gauss} is the copula associated the random variables $\{Y_t\}_{t\in T}$ then we have that the random variables $Z_t = \Phi^{-1}(F_{Y_t}(Y_t)) = \Phi^{-1}(U_t)$ have joint distribution equal to $\mathcal{N}(0,\Sigma)$. This follows from:

$$\begin{aligned} P(Z_1 \leq z_1, \dots, Z_T \leq z_t) &= P(\Phi^{-1}(U_1) \leq z_1, \dots, \Phi^{-1}(U_T) \leq z_T) \\ &= P(U_1 \leq \Phi(z_1), \dots, U_T \leq \Phi(z_T)) \\ &= C_\Sigma^{\text{Gauss}}(\Phi(z_1), \dots, \Phi(z_t)) \\ &= \Phi_\Sigma(z_1, \dots, z_T) \end{aligned}$$

In particular, given the realization $\{y_{t,j}\}_{t\in t, j\in J}$ of the variables $\{Y_t\}_{t\in T}$, an unbiased estimation of the parameter matrix Σ is the empirical covariance matrix $\hat{\Sigma}$ of the samples $\{\Phi^{-1}(\hat{F}_{Y_t}(y_{t,j})\}_{t\in T, j\in J}$, where $\hat{F}_{Y_t}$ is the estimated marginal distribution of the variable Y_t as seen in subsection 1.1.1 and subsection 1.1.2.

Finally, we can generate samples from a Multivariate Gaussian random variable $(Z_t, t \in T)$ having distribution $\mathcal{N}(0,\hat{\Sigma})$. Then the power output scenarios are obtained from these samples by following the previous steps backwards, that is, for each sample, computing $\hat{F}_t^{-1}(\Phi(Z_t))$ for all $t \in T$.

1.2 Generation

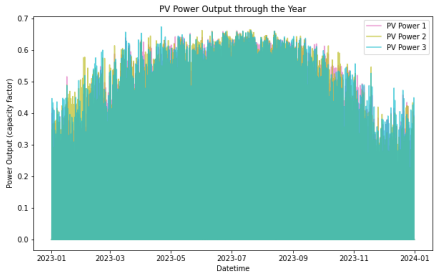
In our project, we used an hourly time step (T=$\{1...8760\}$) and fit the wind and solar distributions separately, thus limiting the computational costs. To fit our model, we used a dataset containing 30 years of data for various European countries, which was collected by [[PS16]].

We observed that the bottleneck of the Scenario Generation algorithm is the Singular Value Decomposition (SVD) of the covariance matrix $\hat{\Sigma}$. Consequently, the computation time changes marginally with the number of generated scenarios d. Thus, in the GUI, we stored the pre-computed SVD matrices for the European countries we worked with, giving the option to rapidly generate a desired amount of scenarios for those countries.

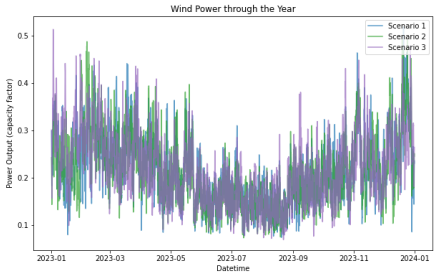
A possible extension could be to also include load scenarios jointly with the generation scenarios through the same approach. This would consider dependence between Energy Demand and weather conditions, but it would necessitate of the historical dataset provided for the corresponding grid, and would also further increase computational costs.

1.2.1 Examples

In this subsection, we present as an example the scenarios generated for Austria. In Figure 1, we can observe the seasonal behavior of power outputs. In Figure 1a, it is evident how solar power is more prominent during summer and less so during winter. This behavior explains why, in the energy grid simulations that will follow, energy is saved in the form of hydrogen during summer and then consumed in winter, when the electric load of urban areas is higher due to heating. Figure 1b shows that wind power also follows a seasonal pattern. In the case of Austria, this is somewhat complementary to solar power, counterbalancing the months of lower solar power output. As we will see in the energy grid simulation, diversity in renewable sources can help diminish the need for energy storage.



(a) Solar Power Output (capacity factor)

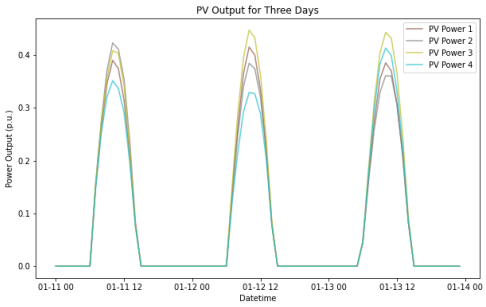


(b) Wind Power Output (capacity factor)

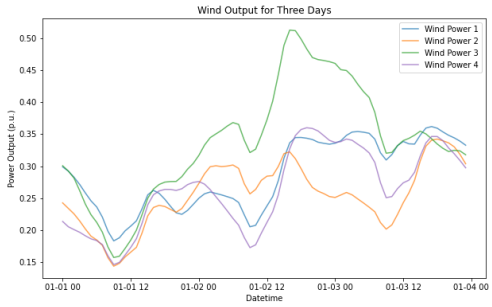
Figure 1: Yearly Power Output of generated scenarios for Austria

In Figure 2 , we observe that both hour to hour and day-to-day interdependence is present in the generated scenarios, as there are instances of chained low solar production values. This accurately mimics the occurrence of cloudy days.

This characteristic is crucial for the yearly Wind Power distribution, as the Electrical Grid needs to have a higher storage capacity to compensate for the lower energy output. It must be noted that a less observed meteorological event in our scenarios, which is of great importance for renewable energy integration, is the Dunkelflaute, also known as Anticyclonic Gloom [[Li+21]]. This event appears to be occurring more frequently due to climate change, implying a higher correlation between Wind and PV production, which could be modeled using the same Copula approach, although the growing size of the matrix Σ would need careful attention to keep the method computationally viable.



(a) Solar Power Output (capacity factor)



(b) Wind Power Output (capacity factor)

Figure 2: 3 days Power Output of generated scenarios for Austria

2 Model Description

The model describes a network in Europe that is to be powered and supplied of hydrogen trough power generated by photovoltaic panels and wind turbines, converted to hydrogen through electrolysis and potentially reconverted in fuel cells. It describes the handling of hydrogen throughout a one year time span, with one hour time steps, in order to meet demand.

The network is represented by an undirected multigraph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ corresponds to the nodes in the network, and $\mathcal{E} = \mathcal{E}_H \cup \mathcal{E}_P$ represents transmission lines ($e \in \mathcal{E}_P$) and hydrogen lines ($e \in \mathcal{E}_H$). Each of these nodes can be in different countries in Europe, and the power generated by wind and solar power depends on the node location. In particular, if node $n \in \mathcal{N}$ is in France, the scenarios for power generation at n will be generated using parameters fitted to France's data.

The operation of the Electric Grid is modeled as a two-stage stochastic program. In the first stage, the capacity expansion of each generator, hydrogen storage, transmission line, and hydrogen pipe is decided. In the second stage, the economic dispatch is solved for each scenario.

2.1 Decision Variables

The main variables that are of interest to the policy maker and are explicitly returned in output are the following:

$$\begin{aligned} \mathrm{ns}_n &: \text{Number of solar units (integer) at node } n \in \mathcal{N}; \\ \mathrm{nw}_n &: \text{Number of wind units (integer) at node } n \in \mathcal{N}; \\ \mathrm{nh}_n &: \text{Storage capacity (kg) (continuous) at node } n \in \mathcal{N}; \end{aligned}$$

Stored hydrogen is considered to be the total of liquid and gas hydrogen to be stored. Our model does not assume a distinction between the two forms, and considers hydrogen to be immediately ready for long-term storage as soon as it is converted from electricity, as well as instantaneously convertible to electricity in fuel cells at need.

To give an indication of how much hydrogen should be kept in gas form at fuel cells, ready to be converted, and how much storage for gas hydrogen should be availlable at electrolyzers to accomodate abundant generation moments, we consider the following decision variables as well (N.B.: they are values *per hour*, to be compared to the time necessary for gas to liquid hydrogen transformation and vice versa):

mhte_n : Maximum hydrogen to electricity capacity (Kg) (continuous), i.e. the$n \in \mathcal{N}$; maximum amount of hydrogen that needs to be converted within a 1h time frame into electricity at node

meth_n : Maximum electricity to hydrogen capacity (MWh) (continuous) at$n \in \mathcal{N}$; node

To accommodate for possible improvements on capacity of existing power lines or hydrogen transport infrastructure, the following variables are added:

$$\begin{aligned} \mathrm{addNTC}_l &: \text{additional net transfer capacity on line } l; \\ \mathrm{addMH}_l &: \text{additional hydrogen transfer capacity on pipe } l. \end{aligned}$$

And parameters for the cost of such new infrastructure have to be given, indexed by edge of the respective graph: $cNTC_l$ and cMH_l.

The second stage variables, indexed by scenario j, time step t, and node $n \in \mathcal{N}$ are:

$$\begin{aligned} \mathrm{H}_{j,t,n} &: \text{Stored hydrogen at time } t, \text{ scenario } j \text{ and node } n \text{ (kg) (continuous)}; \\ \mathrm{HtE}_{j,t,n} &: \text{Hydrogen converted to electricity at time } t, \text{scenario } j \text{ (kg) (continuous)}; \\ \mathrm{EtH}_{j,t,n} &: \text{Electricity converted to hydrogen at time } t, \text{ scenario } j \text{ (MWh) (continuous)}; \end{aligned}$$

Note: all variables are set to be non-negative.

2.2 Parameters

There are a series of parameters that characterize the model and can be modified by the policy maker through the GUI. The main ones are related to capital costs of the infrastructure to be built:

$$\begin{aligned} \mathrm{cs}_n &: \text{Cost of one Solar Panel (€)}; \\ \mathrm{cw}_n &: \text{Cost of one Wind Turbine (€)}; \end{aligned}$$

If not specified through the GUI, the following values are assumed for panels and turbines: cs = 400€, cw = 3 000 000€. In this model we assume no marginal costs for PV and wind power production: the operating costs of the farms throughout their life-cycle can be factored into the capital costs, and there is no additional cost linked to the production itself. We assume the flow of electricity has no marginal cost nor power loss (the modelling of that problem is beyond the scope of this project). Conversely, we do set a cost for the use of hydrogen pipes (or other means of transfer):

cH_edge_l : Cost of transferring 1kg of hydrogen through edge l.

Conversely, the marginal costs of conversion within electrolyzers and power cells are relevant. Thus one can set the following parameters:

chte : Conversion cost of 1 kg of hydrogen to electricity (€/kg);

ceth : Conversion cost of 1 MWh of electricity to hydrogen (€/MWh).

According to the European Hydrogen Market Landcape November 2023 Report [Obs23], "Hydrogen production costs via electrolysis with a direct connection to a renewable energy source in Europe vary from 4.18 to 9.60 €/kg H2 of hydrogen, with the average for all countries being 6.86 €/kg H2". For electrolyzers, we consider the Levelised cost of hydrogen (LCOH) to account for both marginal costs and capital costs. Such cost is dependent on the country's specific market condition and can be calculated through the European Hydrogen Observator tool. Unless specified through the GUI, $ceth$ = 20kg/MWh*10€/kg=200€/MWh is assumed (see the discussion on conversion efficiency below), and $chte$ is assumed to be 2€/kg. Furthermore, the storage of the hydrogen itself has a cost that depends on various factors: capital cost of the technology used for storage, operating costs, length of time that the hydrogen is kept in storage. Thus the following parameters can be set:

ch : Cost of hydrogen storage capacity per unit of hydrogen (€/kg);

ch_t : Cost of storing hydrogen for 1h, per unit of hydrogen (€/(kg·h)).

Thus ch will be simply multiplied by the maximum storage needed (nh), representing capital cost of storage infrastructure, whereas ch_t represents the marginal cost of keeping the hydrogen stored. Unless specified, it is assumed to be $ch = 10$€/kg and $ch_t = 0$€/(kg·h)

Within the electrolyzers and fuel cells, the conversion itself can be more or less efficient:

fhte : efficiency of hydrogen to electricity conversion (scalar between 0 and 1);

feth : efficiency of electricity to hydrogen conversion (scalar between 0 and 1).

It is assumed that 1kg of hydrogen has an energy value of 33kWh. Thus if we consider an electrolyzer operating at maximum efficiency ($feth = 1$), one MWh of electricity yields 1000/33≃30kg of hydrogen. If not specified through the GUI, a value of $feth = 0.66$ is considered, thus 1MWh yields 20kg of hydrogen. Conversely, in a fuel cell operating at maximum efficiency ($fhte = 1$) 1kg of hydrogen yields 33kWh. If not specified through the GUI, a value of $fhte = 0.75$ is considered, yielding 24.75kWh per kg of hydrogen. Actual efficiencies vary a lot depending on the technology used. Furthermore, chemical and physical constraints make it so that efficiencies higher than 0.80-0.85 are currently considered unachievable [DAS20].

Finally, the GUI gives the option to place upper bounds to the variables, based on either technological and physical constraints (dimension of the facilities) or because of political choices (local population unfavourable to wind turbines):

Mns : Maximum number of solar panels that can be installed (integer);
Mnw : Maximum number of wind turbines that can be installed (integer);
Mnh : Maximum hydrogen storage capacity (kg);
Mhte : Upper bound for $mhte$ (kg);
Meth : Upper bound for $meth$ (MWh).

If no bound is given through the GUI, these values will be set to $Mns = 10^6, Mnw = 500, Mnh = 10^9, Mhte = 10^6, Meth = 10^5$). Note: computation time increases significanly when increasing the upper bound for Mnw.

A scenario consists in a different realizations of the following variables, given as input in the model and indexed by scenario j, time step t, and node $n \in \mathcal{N}$:

$\mathrm{ES}_{j,t,n}$: Power output of a single solar panel (MWh)
$\mathrm{EW}_{j,t,n}$: Power output of a single wind turbine (MWh)
$\mathrm{EL}_{j,t,n}$: Electricity load (MWh)
$\mathrm{HL}_{j,t,n}$: Hydrogen load (kg)
$\mathrm{P\_edge}_{j,t,l}$: Power passing through line l during time step t in scenario j (MWh) (continuous);
$\mathrm{H\_edge}_{j,t,l}$: Hydrogen flowing through pipe l during time step t in scenario j (kg) (continuous).

Lastly the following parameters are indexed by edge of the respective graph:

NTC_l : Net Transfer Capacity, that is, maximum amount of electricity that can pass on line l of the electric grid in the span of 1h;
MH_l : Maximum amount of hydrogen that can flow on edge l in 1h.

2.2.1 Objective Function

The cost function is given by the sum of all capital costs of installing infrastructure, all marginal costs of the hour-by-hour hydrogen to electricity and electricity to hydrogen conversions, and minimal costs associated to the variables $mhte$ and $meth$ so that they are minimized through the model. Let $Nnodes$, $NEedges$ and $NHedges$ be the number of nodes, edges on the electric grid graph and

edges of the hydrogen transfer graph respectively. The objective function is modified as follows:

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{\text{Nnodes}} \text{cs}_k \cdot \text{ns}_k + \text{cw}_k \cdot \text{nw}_k + \text{ch}_k \cdot \text{nh}_k + \\
+ \quad & \sum_{l=1}^{\text{NEedges}} \text{cNTC}_l \cdot \text{addNTC}_l + \sum_{l=1}^{\text{NHedges}} \text{cMH}_l \cdot \text{addMH}_l + \\
+ \quad & \frac{1}{d} \sum_{j=1}^{d} \sum_{i=1}^{\text{inst}} \left(\sum_{k=1}^{\text{Nnodes}} (\text{ch\_t}_k \cdot \text{H}_{j,t,k} + \text{chte}_k \cdot \text{HtE}_{j,t,k} + \text{ceth}_k \cdot \text{EtH}_{j,t,k}) + \right. \\
& \left. + \sum_{l=1}^{\text{NHedhes}} (\text{cH\_edge}_l \cdot \text{H\_edge}_{j,t,l}) \right) + \\
+ \quad & \sum_{k=1}^{\text{Nnodes}} 0.01 \cdot (mhte_k + meth_k)
\end{aligned}
$$

The $1/d$ factor in front of the marginal costs allows to average over the scenarios, whereas the capital costs are the same for all scenarios. Thus, ignoring the costs of $mhte$ and $meth$, the objective function value gives an estimate of the actual costs (in €) for the set up and maintenance of the system throughout the length of the scenario (one year).

2.3 Constraints

The following constraints are to ensure that for all time steps t and all scenarios j, the electricity load and the hydrogen load are met. The measure units are MWh and kg respectively, conversion factors are considered for HtE and EtH respecively. Let $Out(n)$ and $In(n)$ indicate the outgoing and incoming edges from node n on the respective graph. Then for each node n, the we have the following flow balance constraints:

$$
\begin{aligned}
\text{Electricity Balance:} \quad & \text{ns}_n \cdot \text{ES}_{j,t,n} + \text{nw}_k \cdot \text{EW}_{j,t,n} + 0.033 \cdot fhte_k \cdot \text{HtE}_{j,t,n} - \text{EL}_{j,t,n} - \text{EtH}_{j,t,n} + \\
& \sum_{l \in Out(n)} \text{P\_edge}_{j,t,l} + \sum_{l \in In(n)} \text{P\_edge}_{j,t,l} \geq 0; \\
\text{Hydrogen Storage:} \quad & \text{H}_{j,t+1,n} = \text{H}_{j,t,n} + 30 \cdot feth_k \cdot \text{EtH}_{j,t,n} - \text{HtE}_{j,t,n} - \text{HL}_{j,t,n} - \\
& - \sum_{l \in Out(n)} \text{H\_edge}_{j,t,l} + \sum_{l \in In(n)} \text{H\_edge}_{j,t,l}
\end{aligned}
$$

We ask that the consumed electricity be less or equal than the produced electricity at all times. On the grid itself, the two sides should be equal, but we observe that $ns \cdot ES_{j,t} + nw \cdot EW_{j,t}$ indicate the maximum power that can be generated with set weather conditions, whereas actual production will be regulated to meet demand through curtailment.

The stored hydrogen at time $t+1$ is the result of what was stored at time t adjusted by what was converted and what was sent to the industrial load. For $t = 24 * 365$ we set the same constraint on hydrogen by considering $t+1$ to be index 1: this way we avoid placing a "start time" at an arbitrary place within the year (time is rendered modulo the year) and we avoid the model asking for conveniently high initial storage values of hydrogen appearing out of thin air.

The total storage and conversion capacities are calculated by minimizing the maximum over time and scenarios of the variables $H_{j,t}, EtH_{j,t}$ and $HtE_{j,t}$, for all scenarios j, time steps t and nodes n:

$$\begin{aligned} \text{Storage Capacity Limit:} &\quad \mathrm{H}_{j,t,n} \leq \mathrm{nh}_n; \\ \text{EtH Conversion Limit:} &\quad \mathrm{EtH}_{j,t,n} \leq \mathrm{meth}_n; \\ \text{HtE Conversion Limit:} &\quad \mathrm{HtE}_{j,t,n} \leq \mathrm{mhte}_n. \end{aligned}$$

Finally, edge capacities on the respective graphs are considered for all scenarios j, time steps t and nodes n:

$$\begin{aligned} \text{Net Transfer Capacity:} &\quad |\mathrm{P\_edge}_{j,t,l}| \leq \mathrm{NTC}_l + \mathrm{addNTC}_l; \\ \text{Hydrogen Transfer Capacity:} &\quad |\mathrm{H\_edge}_{j,t,l}| \leq \mathrm{MH}_l + \mathrm{addMH}_l. \end{aligned}$$

3 Optimization and Time Resolution

The time horizon generated by the scenarios has a time resolution where each time step has a length of one hour. Each value represents the total power (hydrogen) production or demand in the corresponding hour at the node. The smaller the length of each time step, the more accurate the results. However, the number of variables and constraints grows linearly with the number of time steps, making the model intractable (especially in the context of an application) with just a few scenarios.

Moreover, considering every hour in each day of the year is partly redundant, as each day will be similar to neighboring days. Yet, simply considering a sample of days for each season might undermine long-term storage capacity representation.

Given an initial time horizon $\mathcal{T} = \{1, \ldots, T\}$, we can consider partitions of $\mathcal{T}$ as a family of disjoint subsets whose union is $\mathcal{T}$. We only consider those partitions where every subset is an interval of $\mathcal{T}$. We refer to these as time partitions. Given a time partition P, we can consider the corresponding model obtained by considering each interval in P as a single time step. For every I in P, we define:

$$ES_{j,I,n} \coloneqq \sum_{i \in I} ES_{j,i,n}, \quad EW_{j,I,n} \coloneqq \sum_{i \in I} EW_{j,i,n}$$

and similarly for $HL_{j,I,n}$ and $HR_{j,I,n}$. We denote the model obtained by the time partition P as CEP_P.

It is evident that the optimal value of CEP_P is a lower bound for the original problem $CEP_{\mathcal{T}}$, as given a feasible solution $(ns, nw, nh, mhte, meth, H, HtE, EtH, Pedge, Hedge)$ of the latter, we can obtain a solution of the former by taking $(ns, nw, nh, mhte, meth)$ the same as in $CEP_{\mathcal{T}}$ and:

$$Pedge_{j,I,e} = \sum_{i \in I} Pedge_{j,i,e}, \quad Hedge_{j,I,e} = \sum_{i \in I} Hedge_{j,i,e}$$

and similarly for EtH and HtE, and $H_{j,I,n} = H_{j,i_0,n}$ where $I = [i_0, \ldots, i_{|I|}] \in P$. In particular, there is a cost-preserving linear map from the feasible space of $CEP_{\mathcal{T}}$ to the feasible space of CEP_P, making the latter a relaxation of the former.

This is generally true when considering any time partition P' finer than P, where for every $t' \in P'$, there exists $t \in T$ such that $t' \subset t$. In particular, we have the following observation:

Observation 3.1 *Let $V_{\mathcal{P}} \subset \mathbb{R}^{N_{\mathcal{P}}}$ and $V_{P'} \subset \mathbb{R}^{N_{P'}}$ be the space of feasible solutions of $CEP_{\mathcal{P}}$ and $CEP_{P'}$, respectively. There exists a linear map $L_{P'P} : \mathbb{R}^{N_{P'}} \to \mathbb{R}^{N_P}$ such that $L(V_{P'}) \subset V_P$ and $c_P(L(x)) = c_{P'}(x)$, where c_P is the cost function of CEP_P and $c_{P'}$ is the cost function of $CEP_{P'}$.*

Thus, by iteratively solving finer time partitions, we converge to the optimal solution of $\mathcal{P}$.

3.1 Warmstart

Let P' be a finer time partition than P. Given an optimal basic feasible solution of CEP_P, we can use it to efficiently warm start the solution of $CEP_{\mathcal{P}'}$. This is done by, keeping the variables and constraints of CEP_P and adding additional constraints to $CEP_{\mathcal{P}'}$ to further link the additional variables with those of CEP_P. For every $I \in P$, scenario j, and edge k, we add the following constraints to $CEP_{P'}$:

$$\begin{aligned}
\mathrm{EtH}_{j,I,k} &= \sum_{J \in P'(I)} \mathrm{EtH}_{j,J,k} \\
\mathrm{HtE}_{j,I,k} &= \sum_{J \in P'(I)} \mathrm{HtE}_{j,J,k} \\
\mathrm{P_edge}_{j,I,k} &= \sum_{J \in P'(I)} \mathrm{P_edge}_{j,J,k} \\
\mathrm{H_edge}_{j,f,k} &= \sum_{J \in P'(I)} \mathrm{H_edge}_{j,J,k}
\end{aligned}$$

We observe that these constraints imply $H_{j,I,k} = H_{j,J_0,k}$ where J_0 is the leftmost interval of I. This is coherent with the interpretation that $H_{j,I,k}$ corresponds to the initial hydrogen storage in the interval I. This can be seen by summing over J the hydrogen storage constraint

$$\begin{aligned}
0 = &\sum_{i=0}^{|P'(I)|-1} \big(\mathrm{H}_{j,J_{i+1},n} - \ \mathrm{H}_{j,J_i,n} + 30 \cdot feth_k \cdot \mathrm{EtH}_{j,J_i,n} - \mathrm{HtE}_{j,J_i,n} - \mathrm{HL}_{j,J_i,n} \\
&- \sum_{l \in \mathrm{Out}(n)} \mathrm{H_edge}_{j,J_i,l} + \sum_{l \in \mathrm{In}(n)} \mathrm{H_edge}_{j,J_i,l}\big) = \\
&= \mathrm{H}_{j,J_{i+1},n} - \ \mathrm{H}_{j,J_0,n} + 30 \cdot feth_k \cdot \mathrm{EtH}_{j,I,n} - \mathrm{HtE}_{j,I,n} - \mathrm{HL}_{j,I,n} \\
&- \sum_{l \in \mathrm{Out}(n)} \mathrm{H_edge}_{j,I,l} + \sum_{l \in \mathrm{In}(n)} \mathrm{H_edge}_{j,I,l} \\
&= \mathrm{H}_{j,J_{i+1},n} - \ \mathrm{H}_{j,I,n} + 30 \cdot feth_k \cdot \mathrm{EtH}_{j,I,n} - \mathrm{HtE}_{j,I,n} - \mathrm{HL}_{j,I,n} \\
&- \sum_{l \in \mathrm{Out}(n)} \mathrm{H_edge}_{j,I,l} + \sum_{l \in \mathrm{In}(n)} \mathrm{H_edge}_{j,I,l}
\end{aligned}$$

Due to the linearity of the objective function, we do not need to modify the cost vector of CEP_P. Given an optimal basic feasible solution of CEP_P with basis B, we can obtain a basic solution of $CEP_{P'}$ by simply adding the columns corresponding to the variables EtH and HtE. This results in a basic solution of $CEP_{P'}$ with basis B'.

Although this solution is not necessarily feasible for $CEP_{P'}$, the block structure of B' and the unchanged cost vector ensure that the reduced cost vector remains negative. Consequently, we can obtain a dual feasible solution given by $c_{B'}B'^{-1}b$, where b is the constant vector of $CEP - P'$, which can be used as a warm start for the dual simplex algorithm.

4 Results

In this section, we present a series of numerical experiments, which can be reproduced by configuring the network to "small-eu" in the application [Ric23]. The model is solved using Gurobi, with computation times measured on a computer equipped with a 13th Gen Intel(R) Core(TM) i7-13700H processor running at 2.40 GHz and 16 GB of installed RAM.

The example model is a five node network consisting of 5 nodes in different countries: France, Germany, Spain, Italy, and Austria and subject to different wheather and load conditions.

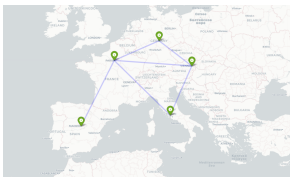


Figure 3: Small-eu network

4.1 Time partition results

While the algorithm described in Section 3 details how successive iterations can be efficiently warm-started, this section provides guidance on selecting the initial time partition and the iteration method. Specifically, the iteration method refers to the strategy used to decide which time intervals in the current partition should be split into smaller intervals.

Regarding the initial time partition, we compare two different strategies against the finest partition $\mathcal{T}$, where each time step is one hour long. This finest partition offers the highest accuracy but is also the most computationally demanding.

The first partition, $\mathcal{T}_1$, groups all days together except for ten evenly spaced days, where each hour is treated as a separate interval. The second partition, $\mathcal{T}_2$, divides each day into two intervals: one for daylight hours and one for night hours. This approach is motivated by the fact that solar power generation occurs only during daylight hours.

In both cases, $\mathcal{T}_1$ and $\mathcal{T}_2$, the finest partition $\mathcal{T}$ is used as the reference for comparison. We observe that there is no particular difference between the two partitions, in fact, surprisingly they yeald the same results and are solved in a similar time (22s), while the model with $\mathcal{T}$ time partition takes a longer time to solve (320s). The main difference between $\mathcal{T}_1$ and $\mathcal{T}_2$ is that the latter seems to better rapresent the daily and seasonal variations of hydrogen storage as can be seen in Figure 4. The difference between the cost of the relaxed models of $\mathcal{T}_1$,$\mathcal{T}_2$ and of $\mathcal{T}$ is of order 10%. Thus, to obtain a better approximation of the optimal solution, we iteratively refine the time partitions.

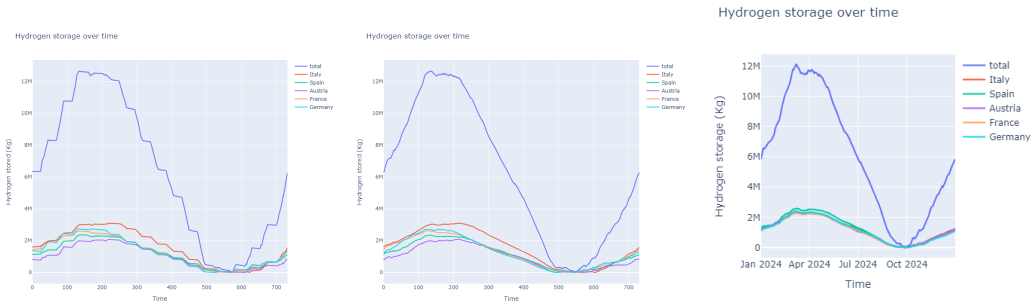


Figure 4: Comparison of hydrogen storage aggregation $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}$.

The iteration method we tested involves randomly selecting a time interval and splitting it into one-hour intervals. While this approach appears to be slightly faster on average than directly optimizing on the final time partition, it can produce suboptimal results when compared to time partitions of similar lengths. To reliably use this method, a validation function needs to be developed to identify which time interval has most violated constraints.

4.2 Analysis of the results

Within a day, we can observe that the stored hydrogen is used to satisfy demand up until dawn: at this point, PV power starts to supply our load and along with generated wind power is enough to refill the depleted hydrogen storage and add to it. In the evening, when the sun sets and electricity load is still high, some of the stored hydrogen is again used to fulfill hydrogen and energy demand. We observe the noticable weekly trend in electricity load, with lower demand on weekends.

When looking at the full year, we can see the seasonal trends in both weather (and thus generation) and demand are depicted in Figure 6. The winter has much higher demand with lower PV power generation, but higher wind power generation. Still, the solution by the solver indicates that it is convenient to store hydrogen during the summer and to use it during winter.

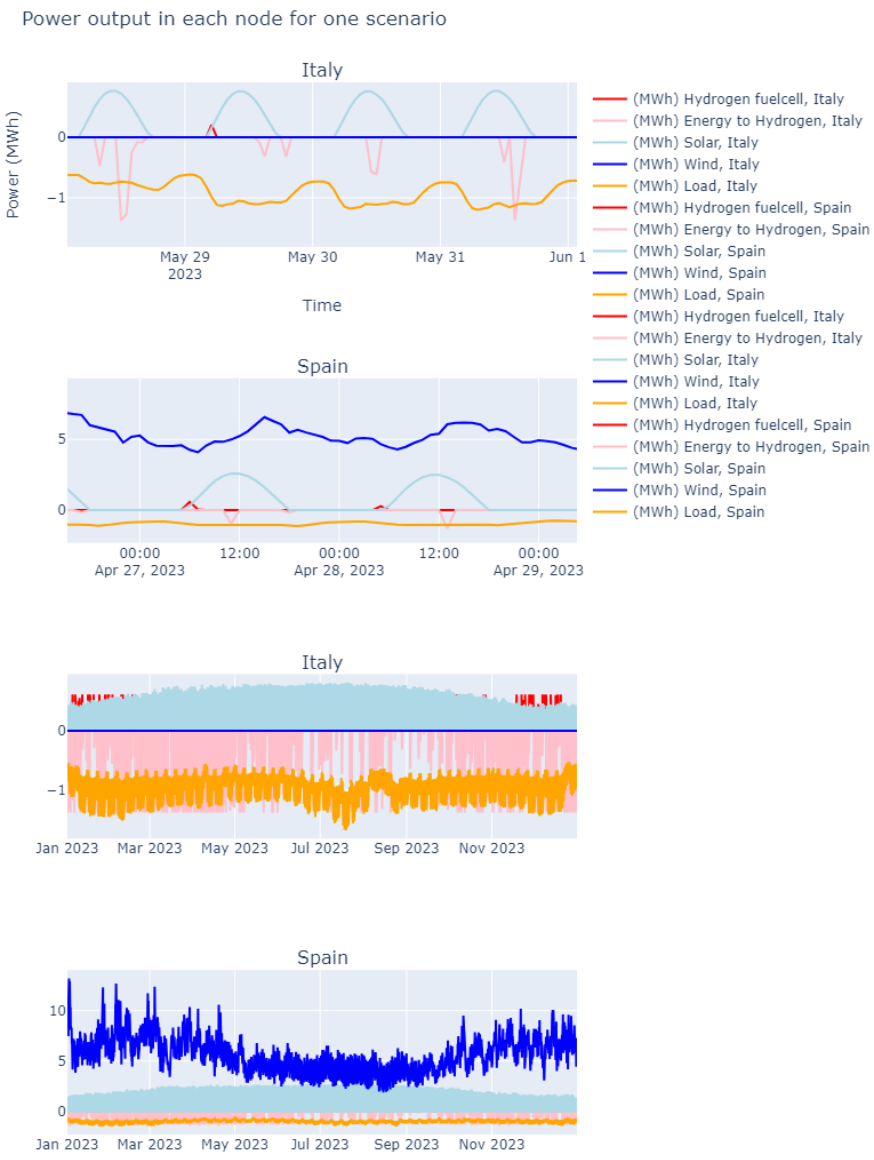


Figure 5: Energy trends in the network, The y-axis represents energy generation and load in MWh.

We observe that Hydrogen efficiency plays a big role in the structure on the network. Low hydrogen efficienty networks tend to have a higher diversity in energy sources to compansate for their variability (as can be seen in 6), since a lot of energy is lost when converting from energy to hydrogen and then back to energy. We observe that the various nodes tend to have one main type of

energy source, depending on the country they are located in. This suggests that high collaboration between countries is required to achieve the best results, but this is only a feasible solution when high amounts of transmission between countries is possible (high NTC values).

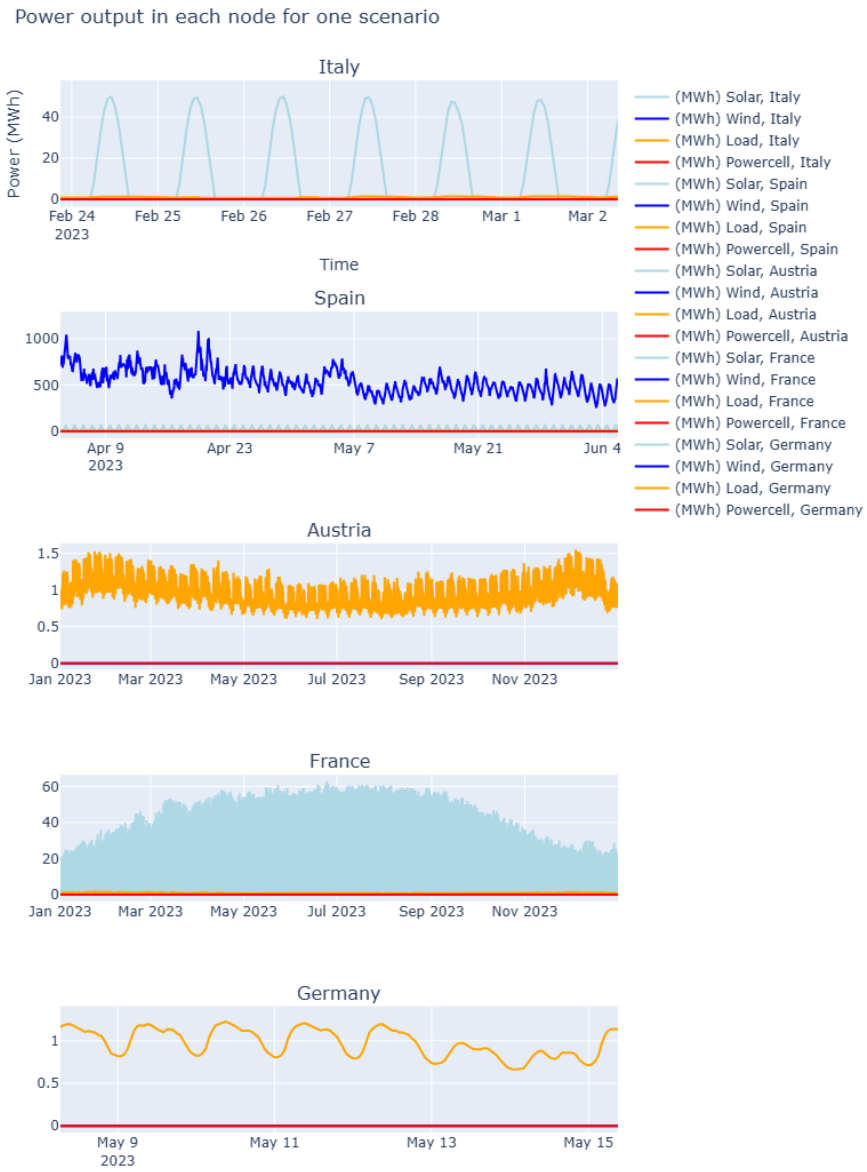


Figure 6: Energy trends in the network, The y-axis represents energy generation and load in MWh.

4.3 Future directions and conclusion.

To enhance the realism and applicability of our energy models, future improvements will focus on integrating wind and solar power scenarios with energy demand models. The parametric modeling of the marginal distributions of solar and wind power provides a robust foundation for incorporating the impacts of climate change. By analyzing shifts in distribution parameters over time, we can adapt our models to better forecast changes in energy production capabilities. Additionally, we plan to extend this approach to account for spatial and generator-type interdependencies. Leveraging real-time data directly from the grid will offer a more nuanced understanding of how various renewable energy sources interact across different timescales and spatial dimensions.

As the number of interdependent variables modeled with the copula approach increases, there is a risk of encountering poor results due to the curse of dimensionality. To address this, one research direction is to approximate the covariance matrix as a banded matrix, thereby reducing the problem's dimensionality. This approach is supported by the assumption that time steps far apart are not significantly correlated. Projection methods can be employed to ensure the matrix remains positive semidefinite. A basic implementation of this was used to save efficiently the covariance matrices.

Another direction for improvement consists in manually constructing the initial solution to warm start the model at each iteration, as Gurobi's warm-start efficiency decreases as the number of scenarios grows. Moreover, while we currently use the same time partition for all scenarios, a more effective approach would involve treating each scenario's time partition independently by identifying the specific time intervals where constraints are violated.

To create a more realistic model, this method could be readily extended to a DC Optimal Power Flow (OPF) framework, though doing so would significantly increase the complexity of the problem. Additionally, from the perspective of policymakers, it would be valuable to incorporate various types of green energy sources into the network. For example, adding hydropower could provide additional energy storage, while geothermal and nuclear energy with their consistent energy supply could reduce the dependence on storage solutions and reduce material consumption.

Bibliography

[DAS20] Furat Dawood, Martin Anda, and George Shafiullah. "Hydrogen production for energy: An overview". In: *International Journal of Hydrogen Energy* 45.7 (2020), pp. 3847–3869. ISSN: 0360-3199. DOI: https://doi.org/10.1016/j.ijhydene.2019.12.059.

[24] *ENTSO-E Power Statistics and Transparency Platform*. Statistical Reports. Data is published based on aggregations of Transparency Platform data, complying with the Transparency regulation as described in the Detailed Data Description document. 2024.

[Hör+18] Jonas Hörsch, Fabian Hofmann, David Schlachtberger, and Tom Brown. "PyPSA-Eur: An open optimisation model of the European transmission system". In: *Energy Strategy Reviews* 22 (2018), pp. 207–215. ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2018.08.012.

[Kha+14] Shahnawaz Farhan Khahro, Kavita Tabbassum, Amir Mahmood Soomro, Lei Dong, and Xiaozhong Liao. "Evaluation of wind power production prospective and Weibull parameter estimation methods for Babaurband, Sindh Pakistan". In: *Energy Conversion and Management* 78 (2014), pp. 956–967. ISSN: 0196-8904. DOI: https://doi.org/10.1016/j.enconman.2013.06.062.

[Li+21] Bowen Li, Sukanta Basu, Simon J. Watson, and Herman W. J. Russchenberg. "A Brief Climatology of Dunkelflaute Events over and Surrounding the North and Baltic Sea Areas". In: *Energies* 14.20 (2021). ISSN: 1996-1073. DOI: 10.3390/en14206508.

[Obs23] European Hydrogen Observatory. "European Hydrogen Market Landscape - November 2023 Report". In: *European Journal of Operational Research* 1 (2023).

[PK09] George Papaefthymiou and Dorota Kurowicka. "Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis". In: *IEEE Transactions on Power Systems* 24.1 (2009), pp. 40–49. DOI: 10.1109/TPWRS.2008.2004728.

[PJ24] Filippo Pecci and Jesse D. Jenkins. *Regularized Benders Decomposition for High Performance Capacity Expansion Models*. 2024. arXiv: 2403.02559 [math.OC].

[PS16] Stefan Pfenninger and Iain Staffell. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". In: *Energy* 114 (2016), pp. 1251–1265. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2016.08.060.

[Ric23] Morgari Riccardi. *MOPTA competition repository*. https://github.com/morgari-b/MOPTA. Accessed: 2024-08-10. 2023.

[Yua+19] Xiaohui Yuan, Chen Chen, Min Jiang, and Yanbin Yuan. "Prediction interval of wind power using parameter optimized Beta distribution based LSTM model". In: *Applied Soft Computing* 82 (2019), p. 105550. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2019.105550.