Optimization and Engineering manuscript No.
(will be inserted by the editor)

Modelling of a fully renewable energy grid with hydrogen storage: time aggregation for a scalable Capacity Expansion Problem

Bianca Urso · **Gabor Riccardi**

Received: date / Accepted: date

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. To adress the variability of wind and solar power output over time, electricity grids expansion plans need to account for multiple scenarios over large time horizons. This significanlty increases the size of the resulting Linear Problem (LP), making them computationally challenging for large scale grids. To tackle this, we propose an approach that aggregates time steps to reduce the problem size, followed by an iterative refinement of the aggregation. Using the previous iteration's solution as a warm start, we introduce methods to select which time intervals to refine at each iteration. The first method employs a validation function, which evaluates with a rolling horizon method the feasibility of the aggregated solutions and selects the time interval on which the validation fail. The second method uses the proportion of net power production in each timestep relative to the aggregated time interval. These selection methods are compared against a random interval selection approach. Lastly, we introduce a class of aggregation methods called structure-preserving aggregations, for which we establish sufficient conditions that ensure the aggregated problem is equivalent to the original within a general framework. This shows the versatility of our approach across different problems. To generate scenarios for wind and photovoltaic (PV) power output, we use scenario generation techniques that account for temporal dependencies. These dependencies are captured using marginal distributions combined with a Gaussian copula, ensuring the generated scenarios accurately reflect the realistic temporal correlations observed in historical data.

B. Urso
IUSS School of Advances Studies, Palazzo del Broletto, Piazza della Vittoria, 15 – 27100 Pavia PV, Italy
Tel: +39 0382 375811
Fax: +39 0382 375899
E-mail: bianca.urso@iusspavia.it

G. Riccardi
Dipartimento di Matematica "F. Casorati" Via Adolfo Ferrata, 5 – 27100 Pavia

Keywords First keyword · Second keyword · need 4-6

Mathematics Subject Classification (2020) 90-10 · 90B15

1 INTRODUCTION

Context; literature overview. Small introduction on LP/MIP. Definition of stochastic optimization, definition of CEP, ED problems.

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 through digital optimization of electric power distribution and transmission in smart grids [?]. One of the main difficulties arising when designing an electric power system relying on renewables is the great variability of the generation of electricity through wind and solar, since these resources are highly dependent on weather conditions. To deal with this variability, a possible solution gaining a lot of traction in recent years is the introduction of an energy storage system relying on hydrogen, converting energy from hydrogen to electricity and vice versa in fuel cells and elecrolyzers. cite source for hydrogen (eg: EU super investing in it). It is of interest to evaluate the optimal solution, in terms of investment plan, to supply the grid along with industrial hydrogen demand in a dependable way. The stochastic nature of the problem though makes it impossible to plan long-term by optimizing on forecasts, and requiring a statistical approach to ensure a robust model. Up to now, common approaches have adopted Stochastic Programming (SP) or Robust Optimization (RO) models, along with hybrid models involving Information Gap Decision Theory or Chance Constraint [1]. While initially favored, the SP approach comes with high computational burden, so RO models have seen more popularity in recent years, despite the drawback of being conservative methods with higher average cost of operation and planning of energy systems.

In the typical setting, the problem to solve is a Capacity Expansion Problem (CEP) regarding infrastructure investments: solar and wind farms, fuel cells, hydrolizers, grid upgrades to augment Net Transfer Capacity (NTC) and so on. Nested within the CEP is an Economic Dispatch (ED) problem concerning the operational costs of said infrastructure. The problem is well suited to be modelled through mixed integer linear programming (MILP), as is explained in detail in [?] and ((other examples?))

The CEP for investment planning requires to look at long time horizons, and on the other hand intra-day variability in generation is the main complexity driver for the ED problem, so the time horizon must be modelled by a large quantity of fairly tight time steps. Furthermore, spacial structure big, even though we ignore it here and just slam everything into country nodes. Thus the temporal and spacial characteristics of the model bring the MILP size to increase rapidly. This is especially demanding in the case of SP, since all the variables from the inner ED problem must be reproduced over all scenarios.

To reduce these costs, one possible approach is to use a Rolling Horizon (RH). The basic technique is described in [?], along with some results regarding quality guarantees for the optimality of the solution. In [?], a rolling horizon approach is used within a RO model to optimize the operation of a micro-grid composed of check serving an isolated area in Chile. other examples of RH used for the ED A similar idea is applied in [?], where integer variables representing capital investments are

initially relaxed and then progressively fixed in successive time steps, reducing the computational costs associated to the search for integer solutions.

A big drawback of the RH approach is that Problem: it is not optimal. In fact, [?]. But better reflects actual decision making based on info that is only available progressively, vs perfect foresigh approach.

Our work aims to build a hybrid model, exploring the use of the RH method as a tool to guide progressively tighter relaxations of the perfect foresight model, rather than as a stand-alone technique. Explain better. Further, RH is then used for the validation of the results for the CEP obtained in the perfect foresight optimization, to ensure that given a solution to the CEP, the ED problem admits feasible solutions even in a limited-foresight environment. The idea is to solve the CEP as to build a grid that can be operated as a "smart grid", through a control system that optimizes on day ahead forecasts.

The model we optimize on in introduce model: components, aim, use Gurobi...

Finally, to generate the scenarios needed to train and test the SP model, a Gaussian Copula method was used. This has been previously done here and here and it's not new right?.

This paper is organized as follows: HOW?

2 MODEL

2.1 LP Formulation

Our 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.

The network is represented by an undirected multigraph $\mathscr{G} = (\mathscr{N}, \mathscr{E})$, where $\mathscr{N}$ corresponds to the nodes in the network, and $\mathscr{E} = \mathscr{E}_H \cup \mathscr{E}_P$ represents transmission lines ($e \in \mathscr{E}_P$) and hydrogen lines ($e \in \mathscr{E}_H$). 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 $n \in \mathscr{N}$ is in France, the scenarios for power generation at n will be generated using parameters fitted to France's data.

Each node has its generators, hydrogen storage, fuel cells and hydrolyzers, for which the capacity is to be decided. Likewise, the CEP aims to solve for transmission line NTC and hydrogen pipe transmission capacity for each edge of the network. The basic formulation of the LP described in this section allows us to solve the CEP with perfect foresight.

We decided to model our problem as a LP problem instead of MILP, as would be standard in the literature SOURCES?. This is because when modelling a large grid with high demand, optimizing on continuous variables and then rounding up to the closest integer for variables *ns* and *nw* notably decreases computational costs, without relevant difference in the final cost. The same wouldn't be possible when optimizing over micro-grids.

The model takes in input the generation and load scenarios of the given area along with various parameters indicating costs and efficiency of the current state of technology and possible upper bounds for the decision variables. The CEP and the ED problems for all scenarios are solved concurrently. When optimizing over multiple scenarios jointly, the solver returns the minimal amount of infrastructure and capacities that is needed to have feasibility (that is, demand met at all time and no blackouts) over all scenarios in input, with minimal cost. Cost is considered to be the sum of capital costs and average operational costs of the infrastructure over the scenarios.

The optimization problem is solved using the Gurobi solver [4].

2.1.1 Decision Variables

The main variables that are of interest to the policy maker are, for each location n, the number of wind turbines and PV panels to install, as well as hydrogen storage capacity. 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. For more detailed models of the management of hydrogen see SOURCES.

Important values to consider when planning the grid are the power capacity and conversion speed of fuel cells and electrolyzers: variables $mhte_n$ and $meth_n$ indicate for each location the maximum amount of energy that can be converted at a single time step respectively from hydrogen to electricity and vice versa. These values are essential to estimate in order to design a grid that can effectively accommodate peak production and supply during low production periods.

explain what NTC is? quote paper. Existing NTC is estimated through the procedure followed by link (page 30), by collecting data from link. Improvements on the existing capacity for power lines or hydrogen transport infrastructure are considered through variables $addNTC_l$ and $addMH_l$. cite EU policy to upgrade grid.

Variables linked to the inner ED problem are indexed by scenario j, time step t, and either node $n \in \mathcal{N}$ or edge $l \in \mathcal{E}_H$ or $l \in \mathcal{E}_P$. For the variables pertaining to transmission on an edge, two distinct variables are considered, one for each direction. This way, all variables are set to be non-negative. This will be relevant for the formulation of the time-aggregated relaxation of the LP problem.

See Table 1 for the summary of all decision variables.

2.1.2 Parameters

There are a series of parameters that describe the grid and are passed to the model. The main ones are related to capital costs of the infrastructure to be built and the following values are assumed for panels and turbines: $cs = 400$€, $cw = 3000000$€. We also set capital costs for fuel cells and electrolyzers: since hydrogen infrastructure is usually obtained by reconverting existing infrastructure from other purposes, the estimation of investment costs is very location dependent and beyond the scope of this work cite article. Thus for our purposes, instead of representing the actual investment

Table 1 Decision variables

Name	Unit	Description
ns_n	-	Number of solar units at node $n \in \mathcal{N}$
nw_n	-	Number of wind units at node $n \in \mathcal{N}$
nh_n	kg	Storage capacity at node $n \in \mathcal{N}$
$mhte_n$	kg	Maximum hydrogen to electricity capacity, at node $n \in \mathcal{N}$
$meth_n$	MWh	Maximum electricity to hydrogen capacity at node $n \in \mathcal{N}$
$addNTC_l$	MWh	Additional net transfer capacity on line $l \in \mathcal{E}_P$
$addMH_l$	kg	Additional hydrogen transfer capacity on pipe $l \in \mathcal{E}_H$
$H_{j,t,n}$	kg	Stored hydrogen at node n, time t, scenario j
$HtE_{j,t,n}$	kg	Hydrogen converted to electricity at time t,scenario j
$EtH_{j,t,n}$	MWh	Electricity converted to hydrogen at time t, scenario j
$P\_edge^{+}_{j,t,l}$	MWh	Power passing through line l at time t, scenario j
$P\_edge^{-}_{j,t,l}$	MWh	Power passing through line l at time t, scenario j
$H\_edge^{+}_{j,t,l}$	kg	Hydrogen transported through line l at time t, scenario j
$H\_edge^{-}_{j,t,l}$	kg	Hydrogen transported through line l at time t, scenario j

for the facilities, a minimal "symbolic" cost is assigned per unit of capacity, so that in minimizing the model estimates needed conversion capacities $mhte_n$ and $meth_n$.

The storage of the hydrogen 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. For our model we only set the parameters *ch*, to be multiplied by the maximum storage needed (*nh*), representing capital cost of storage infrastructure. We ignore marginal costs of keeping the hydrogen stored.

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.

Conversely, the marginal costs of conversion within electrolyzers and power cells are relevant. According to the European Hydrogen Market Landcape November 2023 Report [2], "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.

Parameters $fhte_n$ and $feth_n$ are set as scalars between 0 and 1 to indicate efficiency of the conversion from hydrogen to electricity and vice versa. It is assumed that 1kg of hydrogen has an energy value of 33kWh cite. Thus if we consider an electrolyzer operating at maximum efficiency ($feth = 1$), one MWh of electricity yields 1000/33$\simeq$30kg of hydrogen. For our purposes, a standard 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. We consider

a value of $fhte = 0.75$, 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 [**?**].

Additionally, 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), whereas we do set a cost for the use of hydrogen pipes (or other means of transfer). The existing capacity of transmission lines and hydrogen pipes is also set. mention again the same as before?

Finally, the model allows for upper bounds to be placed on the variables, based on either technological and physical constraints (dimension of the facilities) or because of political choices (e.g. local population unfavourable to wind turbines).

The parameters ES, EW, EL, HL, indexed by scenario, time step and node represent the time series of power generation and load values for different scenarios in every node of the grid. They are generated through the method described in the appendix 6.1

See Table 2 for the summary of all parameters.

2.1.3 Objective Function

The cost function is given by the sum of all capital costs of installing infrastructure, plus all marginal costs of the hydrogen to electricity and electricity to hydrogen conversions and hydrogen transfer on the edges at each time step Let *Nnodes*, *NPedges* and *NHedges* be the number of nodes, edges on the electric grid graph and edges of the hydrogen transfer graph respectively. Let d be the number of scenarios, and T the number of time steps. The objective function is as follows:

$$
\begin{aligned}
\min \quad & \sum_{k\in\mathcal{N}} (\mathrm{cs}_k \cdot \mathrm{ns}_k + \mathrm{cw}_k \cdot \mathrm{nw}_k + \mathrm{ch}_k \cdot \mathrm{nh}_k) + \\
+ \quad & \sum_{k\in\mathcal{N}} (\mathrm{cmhte}_k \cdot \mathrm{mhte}_k + \mathrm{cmeth}_k \cdot \mathrm{meth}_k) + \\
+ \quad & \sum_{l\in\mathscr{E}_P} (\mathrm{cNTC}_l \cdot \mathrm{addNTC}_l) + \sum_{l\in\mathscr{E}_H} (\mathrm{cMH}_l \cdot \mathrm{addMH}_l) + \\
+ \quad & \frac{1}{d} \sum_{j=1}^{d} \sum_{t=1}^{T} \left(\sum_{k\in\mathcal{N}} (\mathrm{ch\_t}_k \cdot \mathrm{H}_{j,t,k} + \mathrm{chte}_k \cdot \mathrm{HtE}_{j,t,k} + \mathrm{ceth}_k \cdot \mathrm{EtH}_{j,t,k}) + \right. \\
& \left. + \sum_{l\in\mathscr{E}_H} (\mathrm{cH\_edge}_l \cdot \mathrm{H\_edge}_{j,t,l}) \right)
\end{aligned}
\tag{1}
$$

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 $mhte_k$ and $meth_k$, 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 time horizon.

Table 2 Model parameters

Name	Unit	Description
cs	€	Cost of one Solar Panel at node n
cw	€	Cost of one Wind Turbine at node n
chte_n	€/kg	Conversion cost of H_2 to electricity
ceth_n	€/MWh	Conversion cost of electricity to hydrogen
ch_n	€/kg	Cost of hydrogen storage capacity
cH_edge_l	€	Cost of transferring 1kg of hydrogen through edge $l \in \mathscr{E}_H$
cNTC_l	€/MWh	Cost of adding NTC to line $l \in \mathscr{E}_P$
cMH_l	€/kg	Cost of adding H_2 transfer capacity to line $l \in \mathscr{E}_H$
cmhte	€/kg	Cost of needed HtE capacity per unit
cmeth	€/MWh	Cost of needed EtH capacity per unit
fhte_n	-	Efficiency of hydrogen to electricity conversion
feth_n	-	Efficiency of electricity to hydrogen conversion
NTC_l	MWh	Net Transfer Capacity on line $l \in \mathscr{E}_P$
MH_l	kg	Hydrogen transfer capacity on edge $l \in \mathscr{E}_H$
Mns_n	-	Maximum number of solar panels installable at node n
Mnw_n	-	Maximum number of wind turbines that can be installed
Mnh_n	kg	Maximum hydrogen storage capacity
Mhte_n	kg	Upper bound for $mhte$
Meth_n	MWh	Upper bound for $meth$
$\text{ES}_{j,t,n}$	MWh	Power output of a single solar panel
$\text{EW}_{j,t,n}$	MWh	Power output of a single wind turbine
$\text{EL}_{j,t,n}$	MWh	Electricity load
$\text{HL}_{j,t,n}$	kg	Hydrogen load

2.1.4 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, and 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. For simplicity, we indicate with $P\_edge_{j,t,l}$ the difference of the corresponding variables $P\_edge^{+}_{j,t,l} - P\_edge^{-}_{j,t,l}$, and analogously for H_edge. When solving, it is sufficient to assign a symbolic cost to them to ensure that only one of them is non zero at each time step. Then for each node n, the following flow balance

constraints are imposed:

$$\begin{aligned}\text{Electricity Balance:}\quad & \mathrm{ns}_n \cdot \mathrm{ES}_{j,t,n} + \mathrm{nw}_k \cdot \mathrm{EW}_{j,t,n} - \mathrm{EL}_{j,t,n} + \\ & + 0.033 \cdot fhte_k \cdot \mathrm{HtE}_{j,t,n} - \mathrm{EtH}_{j,t,n} + \\ & + \sum_{l \in Out(n)} \mathrm{P\_edge}_{j,t,l} + \sum_{l \in In(n)} \mathrm{P\_edge}_{j,t,l} \geq 0; \end{aligned} \tag{2}$$

$$\begin{aligned}\text{Hydrogen Storage:}\quad & \mathrm{H}_{j,t+1,n} = \mathrm{H}_{j,t,n} - \mathrm{HL}_{j,t,n} + \\ & + 30 \cdot feth_k \cdot \mathrm{EtH}_{j,t,n} - \mathrm{HtE}_{j,t,n} + \\ & - \sum_{l \in Out(n)} \mathrm{H\_edge}_{j,t,l} + \sum_{l \in In(n)} \mathrm{H\_edge}_{j,t,l} \end{aligned} \tag{3}$$

We ask that the consumed electricity be less or equal than the produced or received 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. if set to =, see paper??

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 $t = T$ (the last time step) 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:

$$\text{Storage Capacity Limit:}\quad \mathrm{H}_{j,t,n} \leq \mathrm{nh}_n; \tag{4}$$

$$\text{EtH Conversion Limit:}\quad \mathrm{EtH}_{j,t,n} \leq \mathrm{meth}_n; \tag{5}$$

$$\text{HtE Conversion Limit:}\quad \mathrm{HtE}_{j,t,n} \leq \mathrm{mhte}_n. \tag{6}$$

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

$$\text{Net Transfer Capacity:}\quad \mathrm{P\_edge}^{\pm}_{j,t,l} \leq \mathrm{NTC}_l + \mathrm{addNTC}_l; \tag{7}$$

$$\text{Hydrogen Transfer Capacity:}\quad \mathrm{H\_edge}^{\pm}_{j,t,l} \leq \mathrm{MH}_l + \mathrm{addMH}_l. \tag{8}$$

2.2 Validation: Rolling Horizon

While computing the optimal solution on a batch of scenarios by solving the LP model described in section 2.1, the solver "knows the future" for those scenarios. That is, the criteria it uses to determine, at each time step, how much electricity or hydrogen should be converted or transmitted, and where, is a mathematical minimum that is informed by the knowledge of what is needed at any time during the one year scenario. In real life, accurate forecasts for weather, and consequently for

power generation, are known on a day ahead basis, at most two days ahead. We are thus interested in evaluating whether an optimal grid as given by the solution of the CEP problem on a batch of train scenarios can (1) be operated without knowledge of the future in order to satisfy demand on the same scenarios it was trained on, and (2) generalize to new test scenarios.

Let's first consider the case where the grid is composed of a single node. The actions and choices of a power grid administrator of such an isolated micro-grid are very much limited to "given extra energy, store it, up to storage limit". Such a strategy is for example discussed in more detail in [5]. A deterministic system control can be easily designed to check whether a single node CEP solution is sufficient for feasibility over a certain scenario, even without day ahead forecasts. This is not true anymore once the node is taken out of isolation: at each moment, without knowledge of the full future, each fictional administrator at each node must choose whether to store the energy generated at their node, whether to send it to a neighbor (and how much to which neighbor) or how to collect missing energy to match their node's demand.

In order to operate a multi-node grid, we propose to use the Rolling Horizon optimization technique. The basic idea, as described in source is to divide the time horizon into smaller periods and to progressively optimize on each period, passing the variables of the solved periods as fixed to the next. In our case, we are interested in periods of length one day each, representing the knowledge of future that is given to the grid administrator by weather forecasts.

Recall that when optimizing with the LP model described in section 2.1, the solutions to the CEP and to the ED problem for all train scenarios $j \in J_{train}$ are given concurrently. Consider now a test scenario with generation and load time series $ES_{t,n}, EW_{t,n}, EL_{t,n}, HL_{t,n}$ (the test scenario can be in J_{train} or not).

- start with $H_{0,n} = \max_{j \in J_{train}} H_{j,0,n}$
- for each day in the time horizon, optimize the inner ED problem for that day
- set the hydrogen storage levels of the last time step of the day equal to the ones for the first time step of the next day.

The daily ED problem is formulated in a similar way to the LP problem described in section 2.1. The differences are that the CEP variables (ns, nh,...) are fixed, and the hydrogen storage level doesn't loop as it did in the year, but it connects to the following day.

Definition 2.1 (RH-feasibility) Given a solution x_{CEP} to the CEP solved over train scenarios $j \in J_{train}$ and a test scenario j, we consider x_{CEP} to be RH-feasible over scenario j if the RH optimization algorithms terminates at the end of the year and $H_{j,T,n} \geq H_{j,0,n}$ for all nodes $n \in \mathcal{N}$.

We require the last condition to mean that the storage levels at the end of the year are at least as high as they were at the start, to flag as unfeasible solutions that satisfy

demand throughout the year with a net consumption of unproduced hydrogen.

Written as such, the daily optimization would tend to avoid storing hydrogen unless needed within the same day, since operating the elecrolyzers has a cost. This easily renders infeasible scenarios that would be feasible with better storage management. To solve this problem, one can introduce a loss function in the model: for example one can assign a cost to the difference between the hydrogen storage level at time step t and the average over the corresponding variables in the optimal solutions from train scenarios. Thus one defines positive variables $loss_{t,n}$ for each time step t and node n, with posiitive cost, and adds the contraint:

$$loss_{t,n} \geq \frac{1}{d} \sum_{j \in J_{train}} H_{j,t,n} - H^{test}_{t,n} \tag{9}$$

Another option would be to assign a slight negative cost to $H^{test}_{t,n}$, to incentivize filling up the storage, but this can inflate the estimated cost of operating electrolyzers more than necessary.

We observe that the overall solution to the ED problem throughout the full time horizon obtained by means of the RH method is not necessarily optimal, neither with nor without the added loss function. However, some results can be obtained regarding the distance of such solution from the optimal of the ED problem. Indeed: The first step ensures that conditions 1-4 from cite paper are satisfied. Thus we have near optimality guarantee....
For a solution to the ED problem that is closer still to the perfect foresight optimal, more refined RH tecniques can be used. For example, optimizing on two-day forecasts with a daily refresh rate, as done in source can improve the quality of the solution. However, for our purposes, being able to check for feasibility with less than optimal management is sufficient.

3 Time Resolution

The scenarios generated from our gathered data have a time resolution of one hour. where have I said this for the first time?. Such resolution is enough to capture the daily variability of power generation and load. However, the number of variables and constraints grows linearly with the number of time steps, making the model intractable with just a few scenarios. Moreover, when optimizing over a full year, considering every hour of every day 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. Thus we are interested in finding more efficient ways to deal with the time dimension in our problem.

3.1 Time aggregation as model relaxation

We introduce the following concept:

Definition 3.1 Given an initial time horizon $\mathscr{T} = \{1, \dots, T\}$, a time partition $P = \{I_1, ..., I_{T'}\}$ is a partition of $\mathscr{T}$ such that all subsets are intervals. Furthermore, we say that a time partition P' is finer than P if for every $I' \in P'$, there exists some $I \in P$ such that $I' \subset I$.

Given a time partition P, we can consider the problem CEP_P associated to the model obtained by considering each interval in P as a single time step. For every $I \in P$, define:

$$ES_{j,I,n} := \sum_{i \in I} ES_{j,i,n}, \qquad EW_{j,I,n} := \sum_{i \in I} EW_{j,i,n},$$

and analogously for $EL_{j,I,n}$ and $HL_{j,i,n}$.

It is easy to show that the optimal value to the aggregated problem CEP_P is a lower bound for the original problem $CEP_{\mathscr{T}}$. Indeed, given a feasible solution of the latter, we can obtain a solution of the former by fixing the capital infrastructure variables to be the same as in $CEP_{\mathscr{T}}$ and letting all time dependent variables for the inner ED problems be defined as follows:

$$EtH_{j,I,n} = \sum_{i \in I} EtH_{j,i,e}, \qquad HtE_{j,I,e} = \sum_{i \in I} HtE_{j,i,e}, \qquad \Delta H_{j,I,e} = \sum_{i \in I} \Delta H_{j,i,e};$$

similarly, we sum over $P\_edge^{\pm}_{j,I,e}$ and $H\_edge^{\pm}_{j,I,e}$, separately on the two directions. By defining the aggregated variables this way, the aggregated constraints are not violated, thus we get a feasible solution to CEP_P. Observe that all variables with non zero cost are defined as greater or equal to zero, so summing over them we define a cost-preserving linear map from the solution space of $CEP_{\mathscr{T}}$ to the solution space of CEP_P.

The above discussion holds in general in the case of any two partitions P and P' where P' is finer than P. We can summarize the above in the following observation:

Observation 3.2 *Let P and P' be two time partitions of $\mathscr{T} = \{1...T\}$ such that P' is finer than P. Let $V_P \subset \mathbb{R}^{N_P}$ and $V_{P'} \subset \mathbb{R}^{N_{P'}}$ be the spaces of feasible solutions of CEP_P and $CEP_{P'}$, respectively. Then there exists a linear map $L : \mathbb{R}^{N_{P'}} \to \mathbb{R}^{N_P}$ 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 $CEP_{P'}$. Then the optimal solution to CEP_P provides a lower bound for $CEP_{P'}$*

To clarify what time aggregation implies on the model constraints, we express the LP problem in matrix form, in order to highlight its inner structure. The goal is to write our problem in standard form:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} c^T x \\ &\text{s.t.} \quad Ax = b \\ &\qquad x \geq 0 \end{aligned} \tag{10}$$

Fistly, we can reformulate the model described in 2.1 by introducing a new set of variables $h_{j,t,n} := H_{j,t+1,n} - H_{j,t,n}$, replacing the original variables $H_{j,t,n}$. Constraints (3) and (4) shall be reformulated accordingly, with the latter becoming:

$$\sum_{i=1}^{t} h_{i,j,n} \leq nh_n, \quad \forall\, t = 1...T, \tag{4'}$$

In the problem under consideration, we have various types of constraints: Electricity Balance, Hydrogen Balance, Hydrogen Storage, and bounds on the variables. Given a time partition P, we define EXPLAIN constraints in matrix form, then row aggregation

$$\begin{bmatrix} ES_{j,t,n} & EW_{j,t,n} & \dots & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & \dots & 0 & 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \text{ns} \\ \text{nw} \\ \vdots \\ P_l \\ H_l \\ HtE_{j,t,n} \\ EtH_{j,t,n} \\ h_{j,t,n} \end{bmatrix} \geq \begin{bmatrix} EL_{j,t,n} \\ = HL_{j,t,n} \end{bmatrix} \tag{11}$$

A^{CEP} $(ES, EW, \dots)$	0	⋱
0	$\begin{matrix} 1 \\ 1 \\ \vdots \\ 1 \end{matrix}$	$\begin{matrix} -1 & 0 & 0 & 0 & 0 & \dots & 0 \\ -1 & -1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & -1 & -1 & \dots & -1 \end{matrix}$

3.2 Iteration on time partitions

idea: aggregated problem is relaxation, so we can solve it (fast) and use it to warm start another solve with some disaggregated interval. How to choose when to disaggregate? We test 3 methods:

- Random
- validation function
- ro - devise a method to ”measure” how far an aggregated step is from being allowing extension of the solution to a feasible sol

While obtaining a feasible solution to (??) from (??) is not always guaranteed, it is possible under certain assumptions. Here, if B is a matrix having I, J as set of indexes for the rows and columns respctively, then, for all $I' \subset I$ and $J' \subset J$, we denote with $B_{I',J'}$ the submatrix of B havins rows in I' and columns in J'.

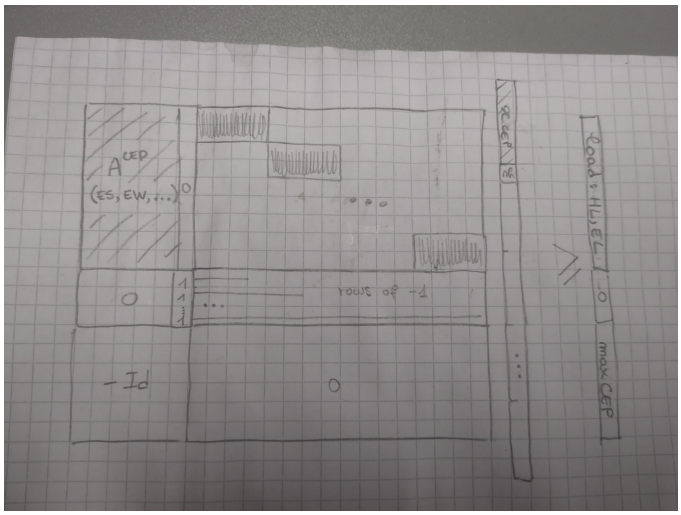

Fig. 1 To be rendered in LaTeX

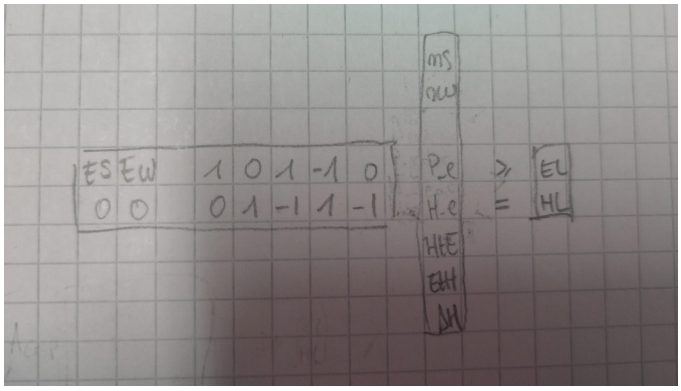

Fig. 2 To be rendered in LaTeX (and correct with coeff 0.033)

Observation 3.3 *If (σ,δ) is a structure-preserving aggregation, let $R \in \sigma$ and $r \in R$. Let $\tilde{x}$ be a solution to the aggregated problem (**??**). If $\tilde{b}_r - \tilde{A}_{R,\delta_{=1}}\tilde{x}_{\delta_{=1}} \neq 0$, define $\rho_r := \frac{b_r - A_{r,\delta_{=1}}\tilde{x}_{\delta_{=1}}}{\tilde{b}_r - \tilde{A}_{R,\delta_{=1}}\tilde{x}_{\delta_{=1}}}$. If $A_{r,\delta_{=1}} = 0$ and $b_r = 0$, then ρ_r can be chosen arbitrarily.*

If $\rho_r \geq 0$ and $x \in \mathbb{R}^n$ satisfies $x_{\delta_{=1}} = \tilde{x}_{\delta_{=1}}$ and $x_{f^r(C)} = \rho_r \tilde{x}_C$ for all $C \in \operatorname{supp}(\tilde{A}_R)_{>1}$, then x satisfies the constraint $A_r x = b_r$ of the original problem.

Proof From the hypothesis and the definition of structure-preserving aggregation, we have:

$$\begin{aligned} A_r x &= \sum_{i \in \text{supp}(A_r)} A_{r,i} x_i \\ &= \sum_{S \in \text{supp}(\tilde{A}_R)_{>1}} A_{r,f^r(S)} x_{f^r(S)} + \sum_{j \in \delta_{=1}} A_{r,j} x_j \\ &= \sum_{S \in \text{supp}(\tilde{A}_R)_{>1}} \tilde{A}_{R,S} \rho_r \tilde{x}_S + \sum_{j \in \delta_{=1}} A_{r,j} x_j \end{aligned}$$

By the definition of ρ_r:

$$\begin{aligned} \rho_r \sum_{S \in \text{supp}(\tilde{A}_R)_{>1}} \tilde{A}_{R,S} \tilde{x}_S &= \rho_r (\tilde{A}_R \tilde{x} - \tilde{A}_{R,\delta_{=1}} \tilde{x}_{\delta_{=1}}) \\ &= \rho_r (\tilde{b}_R - \tilde{A}_{R,\delta_{=1}} \tilde{x}_{\delta_{=1}}) \\ &= b_r - A_{r,\delta_{=1}} \tilde{x}_{\delta_{=1}} \end{aligned}$$

Thus, we obtain:

$$A_r x = b_r$$

□

A structure-preserving aggregation does not inherently guarantee feasibility for all constraints in the original problem. However, Observation 3.3 illustrates how to partially reconstruct a solution x for a specific constraint r by appropriately scaling the aggregated variables within the support of A_r.

We can now define the hypergraph associated to the aggregation (σ, δ).

Definition 3.4 The *hypergraph associated to the aggregation* (σ, δ) is the hypergraph $\mathcal{N}, \mathcal{E}$ having as nodes the aggregated variables $\mathcal{N} := \delta_{>1}$ and as edges the subsest of $\mathcal{N}$ that appear together in not row-agnostic constraints.

When two edges (constraints) in the hypergraph, r and r', share aggregated variables, the scaling factors ρ_r and $\rho_{r'}$ must be equal to maintain consistency. Then if ρ_r can be defined consistently, by applying observation 3.3, to all $r \in R \in \sigma_{>1}$ we can construct a feasible solution for the unaggregated problem (**??**). Thus we have:

Proposition 3.5 *If (σ, δ) is a structure-preserving aggregation. Let $\tilde{x}$ be a solution to the aggregated problem* (**??**)*. If for all $r \in R \in \sigma_{>1}$ such that $\tilde{b}_r - \tilde{A}_{R,\delta_{=1}} \tilde{x}_{\delta_{=1}} = 0$ we have $A_{r,\delta_{=1}} = 0$ and $b_r = 0$. Define $\rho_r := \frac{b_r - A_{r,\delta_{=1}} \tilde{x}_{\delta_{=1}}}{\tilde{b}_R - \tilde{A}_{R,\delta_{=1}} \tilde{x}_{\delta_{=1}}}$ for all $r \in R \in \sigma_{>1}$ such that $\tilde{b}_r - \tilde{A}_{R,\delta_{=1}} \tilde{x}_{\delta_{=1}} \neq 0$. If $\rho_r \geq 0$ and is constant over the connected components of the hypergraph associated to (σ, δ). Then $x_{\delta_{=1}} := \tilde{x}_{\delta_{=1}}$ and $x_{f^r(C)} := \rho_r \tilde{x}_C$ for all $C \in \text{supp}(\tilde{A}_R)$ and $C \in \delta_{>1}$ is well defined and is feasible solution for the unaggregated problem* (**??**).

While row aggregation of a linear problem is a relaxation of the original problem, the same does not apply to column aggregation. However, the column aggregation used for the Capacity Expansion Problem in this work is still a relaxation. In general a column aggregation of a linear problem is a relaxation of the original problem whenever it is a *constant-coefficients column aggregation*, that is:

Definition 3.6 A column aggregation of a linear problem with weights ω_c, $c \in C \in \delta$ is a *constant-coefficients column aggregation* if for all $c \in C \in \delta$ for all rows $r \in [m]$, we have $\omega_c A_{r,c} = \frac{1}{|C|}$ or $\omega_c A_{r,c} = 0$.

That is if for every set $C \in \sigma$ of variables that are aggregated together, each variable in C has the same coefficients in every row of the aggregated problem defined by σ and the coefficient is equal to the aggregation weight of the corresponding variable, except for those rows where all the coefficients of the variables are zero. We then substitute the columns corresponding to C with a vector containing one in every row in which the coefficients in C are non zero, otherwise we substitute with zero.

Proposition 3.7 *If (σ, δ) is a structure-preserving, constant-coefficients column aggregation, if the hypothesis of proposition 3.5 holds then the aggregated problem* (**??**) *is exact and x is an optimal solution.*

Proof Since for observation **??**, the cost of the aggregated problem is equal to the cost of x in the unaggregated problem, we only need to show that the aggregated problem is a relaxation of the unaggregated problem. Let x be a solution to the unaggregated problem (**??**). For all $C \in \delta_{=1}$ let $\tilde{x}_C := x_C$. Since $\{f^r(C)\}_{r \in \cup_{R \in \sigma_{>1}}, C \in \delta} = \cup_{C \in \delta_{>1}} C$, for all $c \in C \in \delta_{>1}$, exists $r \in R \in \sigma_{>1}$ and $C \in \delta_{>1}$ such that $f^r(C) = c$. Then let $\tilde{x}_C := \sum_{c \in C} A_{r,c} x_c$. Since if $f^r(C) = f^{r'}(C')$ implies that $C = C'$, x is well defined. Lastly for all $R \in \sigma$ we have:

$$\tilde{A}_R \tilde{x} = \sum_{r \in R} \left(\sum_{C \in \delta_{=1}} \omega_r A_{r,c} \tilde{x}_C + \sum_{C \in \mathrm{supp}(\tilde{A}_r)_{>1}} \tilde{x}_C \right) = \sum_{r \in R} \left(\sum_{C \in \delta_{=1}} \omega_r A_{r,c} x_C + \sum_{C \in \mathrm{supp}(\tilde{A}_r)_{>1}} \sum_{c \in C} A_{r,c} x_c \right) = \sum_{r \in R} \tilde{\rho}_r b_r = \tilde{b}_R$$

□

3.3 Application to Capacity Expansion Problem

We now apply the results of the previous section to the Capacity Expansion Problem.

It can be easily checked that:

Observation 3.8 *The aggregation of the Capacity Expansion Problem (CEP) is a structure-preserving aggregation with constant coefficients. Furthermore, the constraints* (4′) *are ρ-agnostic.*

Since the only constraints linking variables across different time steps are those in (4′), the connected components of the hypergraph representing the aggregation correspond to the hypergraphs of the Energy Dispatch (ED) at each time step. Specifically, there is one connected component per time step, with nodes representing the variables at that step. To construct a feasible solution from the aggregated problem, it is sufficient for ρ_r to be constant across constraints within the same time step.

Using the definition of ρ_r, it can be shown that for each time step $t \in T \in \mathcal{T}$, this corresponds to the ratio between the net energy production at time t and the net energy production over the interval T. Hence, the conditions of Proposition 3.5 are

met if, for a fixed time step t, this ratio is the same for all nodes in the network, and if, during the interval T, the net production at each node is consistently positive or consistently negative. Thus, we state:

Observation 3.9 *If for each time step $t \in T \in \mathscr{T}$, the ratio of the net energy production at time t to the net energy production over the interval T is the same for all nodes in the network, and if, throughout the interval T, the net production at each node is either always positive or always negative, then the aggregated problem is exact.*

Although this is not generally the case, it suggests an iterative procedure for refining the solution of the aggregated problem. At each iteration, we refine the time partition of the aggregated problem, selecting the time interval to be refined based on the extent to which it violates the conditions of Observation 3.9. Specifically, we split intervals where the net production changes sign, and those intervals T in which, for a fixed time step $t \in T$, the ratio between the net energy production at time t and the net energy production over the interval T exhibits the largest variance.

4 COMPUTATIONAL RESULTS

To evaluate the methodologies presented in this paper, we examine a 5-node network over a one-year span, utilizing timesteps of 1 hour across two distinct scenarios. In this section, we compare three approaches for iterating on the aggregated problem: (1) randomly selecting the interval for refinement, (2) selecting the interval with the highest ρ-variance as defined in Section **??**, and (3) selecting the interval where the validation function **??** fails. The scenarios are generated as outlined in Section **??** in the Appendix. The computational tests were conducted on an Intel(R) Core(TM) i7-13700H CPU @ 2.40GHz with 16 GB of RAM.

Plot 3 illustrates the cost variation at each iteration using the ρ selection method compared to random interval selection. The ρ selection method demonstrates a faster increase in cost increase than the random selection method, with comparable optimization times: 174 seconds for the ρ selection method and 155 seconds for the random selection method over 10 iterations. In plot 4, we compare the cost variation at each iteration using the validation function for interval refinement against the random interval selection method. The results indicate that the former method yields a significantly faster increase, implying quicker convergence to the optimal solution. However, this approach incurs greater computational time: many seconds s. It is worth noting that while both the ρ and random iteration methods continue even upon encountering a feasible solut ion for the original problem, the validation function iteration may halt before reaching the maximum iteration limit if the current solution is feasible, thereby being optimal for the original problem.

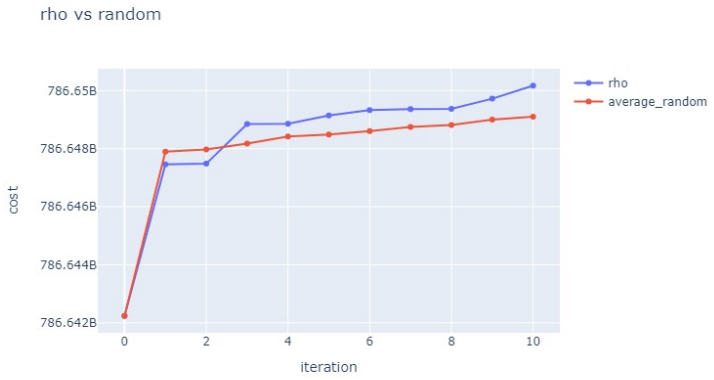

Fig. 3 Cost variation at each iteration using the ρ selection method compared to random interval selection.

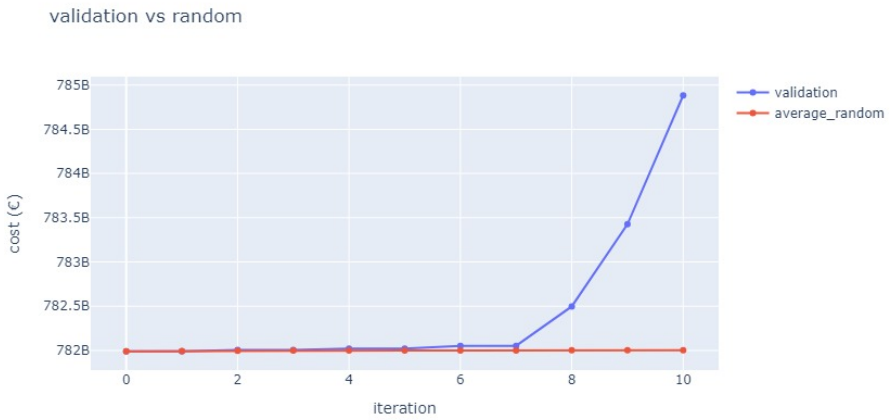

Fig. 4 Cost variation at each iteration using the validation function for interval refinement against the random interval selection method.

5 CONCLUSION

6 APPENDIX

6.1 SCENARIO GENERATION

To estimate the optimal capacities for the CEP through a stochastic approach, realistic and diverse weather scenarios are needed, so to capture the variability and uncertainty of power generation through renewable sources over extended periods. In order to generate such scenarios, samples are extracted from a joint probability density function (PDF) fit on historical data. In the following subsection, we use Y_t to

denote the stochastic process of generated power observations for either solar or wind in a single country. In our project, we used an hourly time step (T={1...8760}) and fit the wind and solar distributions separately for each country considered. To model the marginal probability distributions corresponding to the power output of wind turbines for each hour of the year, a Weibull distribution was used, justified by its proven effectiveness in capturing the variability and skewness of wind power distributions [?]. For solar power, Beta distributions were employed, as in [?].

To fit our model, we used a dataset containing 30 years of data for various European countries, which was collected by [3]. On the other hand, electricity load is taken from the ENTSO-E Statistical Reports. explain how is HL obtained In this simple model, while fitting on historical data we did not account for possible changes in future climate, since the focus lies mostly in the computational aspect.
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: The Gaussian Copula represents well the coupled behavior in renewable stochastic systems [?].

A possible improvement of the generation process could be to fit wind and PV data jointly in the copula step, potentially also including load scenarios 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.

6.1.1 Parametric Estimation of Wind Power distribution

The parameters defining the Weibull Distribution are estimated using the Maximum Likelyhood Estimation (MLE). 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 :

$$\frac{\partial \log L}{\partial \theta} = -\frac{n\gamma}{\theta} + \frac{\gamma}{\theta^2}\sum_{i=1}^{n} x_i^{\gamma} = 0 \tag{12}$$

Eliminating θ, we get:

$$\left[\frac{\sum_{i=1}^{n} x_i^{\gamma} \log x_i}{\sum_{i=1}^{n} x_i^{\gamma}} - \frac{1}{\gamma}\right] = \frac{1}{n}\sum_{i=1}^{n} \log x_i \tag{13}$$

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:

$$\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^{\hat{\gamma}} \right)^{\frac{1}{\hat{\gamma}}} \tag{14}$$

6.1.2 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:

$$\begin{cases} \alpha = \mathbb{E}[X] \left(\frac{\mathbb{E}[X](1-\mathbb{E}[X])}{\text{Var}[X]} - 1 \right) \\ \beta = (1-\mathbb{E}[X]) \left(\frac{\mathbb{E}[X](1-\mathbb{E}[X])}{\text{Var}[X]} - 1 \right) \end{cases} \tag{15}$$

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

6.1.3 Parametric Copula Estimation

The cumulative density function of both the Weibull and Beta distributions are continuous and invertible. Therefore, the random variables $U_t := F_{Y_t}(Y_t)$ 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

$$C(F_{Y_1}(y_1), \ldots, F_{Y_T}(y_{|T|})) = P(Y_1 \leq y_1, \ldots, Y_{|T|} \leq y_{|T|}). \tag{16}$$

This function always exists because of Sklar's Theorem cite Sklar?. For a given correlation matrix Σ, the Gaussian Copula with parameter matrix Σ is defined as

$$C_\Sigma^{\text{Gauss}}(u_1, \ldots, u_T) := \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_T)),$$

where Φ, Φ_Σ are the cumulative distribution functions of Gaussian variables having distribution $\mathcal{N}(0,1)$ and $\mathcal{N}(\mathbf{0}, \Sigma)$ respectively. In particular if C_Σ^{Gauss} is the copula associated with 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:

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

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 6.1.1 and subsection 6.1.2.

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 $t \in T$.

References

1. Michal Jasinski, Arsalan Najafi, et al, Operation and Planning of Energy Hubs Under Uncertainty - A Review of Mathematical Optimization Approaches, 2022
2. Author, European Hydrogen Market Landscape - November 2023 Report, page numbers. Publisher, place (year)
3. Stefan Pfenninger and Iain Staffell. “Long-term patterns of European PV output us- ing 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.
4. Gurobi Optimization, LLC: Gurobi Optimizer Reference Manual (2024). https://www.gurobi.com
5. M. Hashem Nehrir, Power Management of a Stand-Alone Wind/Photovoltaic/Fuel Cell Energy System, IEEE Transactions on Energy Conversion, 2008. DOI: 10.1109/TEC.2007.914200