

# Representative Days and Hours with Piecewise Linear Transitions for Power System Planning

Mojtaba Moradi-Sepahvand  
Department of Electrical Sustainable Energy  
Delft University of Technology  
Delft, The Netherlands  
m.moradisepahvand@tudelft.nl

Simon H. Tindemans  
Department of Electrical Sustainable Energy  
Delft University of Technology  
Delft, The Netherlands  
s.h.tindemans@tudelft.nl

**Abstract**—Electric demand and renewable power are highly variable, and the solution of a planning model relies on capturing this variability. This paper proposes a hybrid multi-area method that effectively captures both the intraday and interday chronology of real data considering extreme values, using a limited number of representative days, and time points within each day. An optimization-based representative extraction method is proposed to improve intraday chronology capturing. It ensures higher precision in preserving data chronology and extreme values than hierarchical clustering methods. The proposed method is based on a piecewise linear demand and supply representation, which reduces approximation errors compared to the traditional piecewise constant formulation. Additionally, sequentially linked day blocks with identical representatives, created through a mapping process, are employed for interday chronology capturing. To evaluate the efficiency of the proposed method, a comprehensive expansion co-planning model is developed, including transmission lines, energy storage systems, and wind farms.

**Index Terms**—Energy Storage Systems, Piecewise Linear Optimal Power Flow, Renewable Energy Sources, Time Series Aggregation, Transmission Expansion Planning.

## I. INTRODUCTION

In modern power systems, capturing variability in electrical load and renewable energy source (RES) outputs during operational periods is vital for effective power system planning. However, integrating short-term operational variability into long-term studies presents a substantial challenge. Attempting to account for every hourly time period across multiple years results in large intractable problems. Therefore, to accurately represent the variability in load and RES output power, the extraction of suitable representative time periods becomes necessary. Commonly employed in the literature [1]–[4], extracting representative days (RDs) using clustering algorithms has been a prevailing method. These RDs, each spanning 24 hours, must faithfully maintain the chronology and extreme values of the underlying time series. Notably, when long-term energy storage systems (ESSs) are involved, interday chronology preservation becomes essential [1].

Past research, such as [2], employed hierarchical clustering to extract typical RDs for weekends, weekdays, and holidays in each season. Spectral clustering, as presented in [3], was utilized for RD extraction by considering net load and the ramp of net load duration curves. However, both approaches lacked consideration for interday chronology and capturing extreme values. The study by [4] introduced an RD selection method focused on capturing extreme operating conditions for power system planning while disregarding interday chronology. Additionally, [5] employed a time series aggregation (TSA) method, relying on system state extraction, to maintain chronological data sequence while disregarding extreme values. Notably, the considerable number of sequential hourly system states in [5] does not significantly decrease the problem complexity [6].

Efforts to capture data chronology introduced the chronological time period clustering (CTPC) method in [7]. In addition, this method’s inability to preserve extreme values was addressed and improved in [8]. However, CTPC-based methods remained inadequate for long-term ESS cycle modeling [1]. To address these challenges, [1] developed a hybrid clustering-based algorithm. It effectively maintained chronology and extreme values by introducing sequentially linked day blocks (SLDs) for long-term ESS modeling. According to comprehensive reviews of TSA approaches presented in [6], [9], many papers employed RDs with 24-hour periods, applying the same data patterns for all system buses [1]–[5], [7], [8]. Despite this, further complexity reduction can be achieved by extracting appropriate representative hours (RHs) or time points (RTPs) within each RD. The complexity challenge is highlighted by the fact that power systems are diverse, exhibiting distinct load and RES generation patterns across multiple areas. Consequently, adopting identical load and RES generation patterns for all buses yields unrealistic results.

Conventional power system optimal power flow (OPF) models implicitly utilize piecewise constant (PWC) formulations, describing injections and loads with average power levels for a time interval (effectively energy). However, PWC models tend to overestimate flexibility, as they inaccurately model ramping and reserves [10]. In contrast, piecewise linear (PWL) OPF formulations (also called ‘power-based’) offer improved operational flexibility by more accurately representing in-

---

This work is part of the project Heuristic Efficient Proxy-based Planning of Integrated Energy Systems (HEPPIE), funded by Réseau de Transport d’Électricité (RTE).

stantaneous power trajectories [11], [12]. In [10], a power-based (i.e., PWL) economic dispatch model is introduced, and [12] presents a security-constrained unit commitment power-based formulation, both offering superior accuracy compared to traditional energy-based models. Similarly, the power-based generation expansion planning model in [11] demonstrates enhanced accuracy in representing flexibility requirements while considering short-term ESSs. However, this model neglects long-term ESS and multi-area data considerations, as well as inter-period chronology and the preservation of extreme values.

To tackle these limitations, this paper introduces a multi-area method that effectively captures intraday and interday chronology, while considering extreme values. The proposed method achieves this using a limited number of RDs and RTPs within each RD, as well as an adaptive optimization-based procedure to improve intraday chronology and extreme values capturing by allocating RTPs across RDs in accordance with the complexity of each day. Furthermore, in order to capture the interday chronology necessary for modeling long-term ESSs, the proposed method builds on the ERD methodology [6] in combination with the SLD process to merge adjacent days introduced in [1]. We further enhance our method by transforming the problem formulation from a traditional PWC to a PWL formulation. This enhancement contributes to improved modeling accuracy by better representing operational flexibility. To validate the proposed method effectiveness, an expansion co-planning model is developed that includes transmission lines, both short (battery) and long-term (hydropower pumped) ESSs, and wind farms. In what follows, the main contributions are outlined:

- Proposing an optimization-based PWL RTP extraction method to improve intraday chronology and extreme value capturing by extracting both equal and unequal numbers of RTPs within each RD.
- Improving generation cost modeling for thermal units within the PWL framework.
- Developing a multi-area PWL-adapted clustering-based algorithm for extreme-value sensitive RD extraction.

## II. REDUCTION OF DAYS AND TIME POINTS

### A. Piecewise Linear Model

In this paper, we use PWL OPF formulations in the proposed co-planning model. This offers improved modeling of ramping and spinning reserve in comparison with traditional PWC models. More importantly, it is anticipated that these benefits *grow* as the model representation becomes increasingly coarse-grained, i.e., when fewer RTPs are used.

The shift from PWC to PWL necessitates a careful description of time in an optimization model. In the PWC representation, measurements refer to an *interval* of time, whereas in the PWL representation, values refer to an *instant* in time.

In the following, the index  $t$  is an integer time coordinate that refers to a point in time, and a pair  $(t-1, t)$  refers to a time

interval. To give a specific example of the PWL interpretation, snapshot power measurements  $P_t$  can be thought to represent a *continuous* function of  $t'$ :

$$P(t') = P_t + (t' - t)(P_{t+1} - P_t) \quad \text{for } t' \in [t, t+1] \quad (1)$$

The energy consumed/produced during the interval  $[t, t+1]$  is

$$E(P_t, P_{t+1}) = \frac{1}{2} \Delta_t (P_t + P_{t+1}) \quad (2)$$

where  $\Delta_t$  is the (real) time between time points indexed by  $t$  and  $t+1$ .

### B. Representative Days

In this paper, the extreme-value sensitive method in [1] is developed as a multi-area clustering-based algorithm to select RDs. Input data consisting of hourly measurements of multiple features (wind, load) in multiple areas, was first divided into daily sequences consisting of 25 hourly measurements (with midnight being represented in both adjacent days). Euclidean distance between days (incl. all features and areas) was used to hierarchically cluster days until the target number was reached. By default, days during which one area's maximum net load (load minus maximum available wind capacity) occurs, are marked as *extreme days*. These days are preserved, to improve planning for system adequacy.

After extracting RDs, the SLDs are created through a mapping process according to [1]. The outputs of the developed RD extraction method are: sets of unordered RDs, i.e.,  $\mathcal{D}$ , and ordered SLDs, i.e.,  $\mathcal{SD}$ , the weight of each RD  $d$ , i.e.,  $\omega_d$ , the association matrix  $D_{sd,d}$  between RD  $d$  and SLD  $sd$ , the number  $n_{sd}^B$  of RD repetitions within an SLD  $sd$ , and instant load and wind representative factors in each area  $a$ , RD  $d$  and time  $t$ , i.e.,  $FL_{a,d,t}$  and  $FW_{a,d,t}$ .

### C. Sparse Day Representation

In the PWL modeling framework, a representative day is initially represented by 25 hourly time points (midnight to midnight). In this section, we describe an optimization-based method to reduce this number with minimum loss of modeling accuracy. Two variations of the method are proposed: one in which a specified number of time points is extracted from a single RD, and another in which a total number of time points is specified for all RDs. The latter makes it possible to extract unequal numbers of RTPs within each RD, allowing for the extraction of more RTPs from RDs with higher hourly variation. The optimization for a single day is given by

$$O(y_{a,t,f}^{RD}, \bar{r}) = \min_{Z_{a,t,f}, I_t, ER_{a,t,f}^{+,-}} \sum_{a,t,f} ER_{a,t,f}^+ + ER_{a,t,f}^- \quad (3a)$$

$$ER_{a,t,f}^+ - ER_{a,t,f}^- = y_{a,t,f}^{RD} - Z_{a,t,f} \quad \forall a, t, f \quad (3b)$$

$$ER_{a,t,f}^+ \geq 0, \quad ER_{a,t,f}^- \geq 0, \quad \forall a, t, f \quad (3c)$$

$$I_t \in \{0, 1\} \quad \forall t \quad (3d)$$

$$\sum_t I_t = \bar{r}, \quad I_0 = I_{24} = 1 \quad (3e)$$

$$Z_{a,t,f} \leq y_{a,t,f}^{RD} + M(1 - I_t) \quad \forall a, t, f \quad (3f)$$

$$Z_{a,t,f} \geq y_{a,t,f}^{RD} - M(1 - I_t) \quad \forall a, t, f \quad (3g)$$

$$Z_{a,t,f} \leq 1/2(Z_{a,t-1,f} + Z_{a,t+1,f}) + MI_t \quad \forall a, t, f \quad (3h)$$

$$Z_{a,t,f} \geq 1/2(Z_{a,t-1,f} + Z_{a,t+1,f}) - MI_t \quad \forall a, t, f \quad (3i)$$

Here  $a$ ,  $t$ , and  $f$  are indices of area, time points, and feature vectors.  $y_{a,t,f}^{RD}$  are the values of the relevant feature vector (load, wind) for the given RD  $d$ . For  $f = \text{wind}$ ,  $y_{a,t,f}^{RD} = FW_{a,d,t}$ ; for  $f = \text{load}$ ,  $y_{a,t,f}^{RD} = FL_{a,d,t}$ . The objective (3a) is to minimize the integrated absolute error between  $y_{a,t,f}^{RD}$  and its approximation  $Z_{a,t,f}$ , summed over all areas, time points, and features. Positive and negative deviations are accumulated in  $ER_{a,t,f}^+$  and  $ER_{a,t,f}^-$ , respectively ((3b)-(3c)). The sparse representation is constructed by selecting  $\bar{r}$  time points using binary variables  $I_t$ , where the first and last time of the day are always selected (3e). Constraints (3f)-(3g) indicate that if a time point is selected (i.e.,  $I_t = 1$ ),  $Z_{a,t,f}$  takes the value of  $y_{a,t,f}^{RD}$  (a big- $M$  formulation is used). Otherwise, it takes the average of the neighboring values ((3h)-(3i)); this is not required for  $t = 0$  and  $t = 24$ .

The algorithm to balance RTPs across RDs minimizes the maximum error across RDs using a greedy algorithm. A flowchart for the algorithm is depicted in Fig. 1. First, a minimum number of RTPs ( $r_d = \bar{r}_{min}$ ) is assigned to each RD. After solving the optimization (3) for all RDs, the day with the largest mismatch is identified and the number of RTPs for that day is increased by one. The algorithm continues until the total number of RTPs is equal to the total number of RDs ( $|\mathcal{D}|$ ) times the desired average number of RTPs ( $r_{avg}$ ) per RD.

After termination of the RTP selection algorithm, only the  $r_d$  selected RTPs within each day are extracted for further use. Let  $J_{d,k}$  be the index (hour) of the  $k$ -th selected RTP on day  $d$ , with  $J_{d,0} = 0$  and  $J_{d,r_d-1} = 24$ . For consistency of notation between models, we let index  $t$  refer to the selected RTPs, so e.g.,  $FW_{a,d,t} \leftarrow FW_{a,d,J_{d,t}}$ . Moreover, to account for unequal time steps, we define

$$\Delta_{d,t} = J_{d,t+1} - J_{d,t}, \quad \forall d, t \in \{0, \dots, r_d - 1\} \quad (4)$$

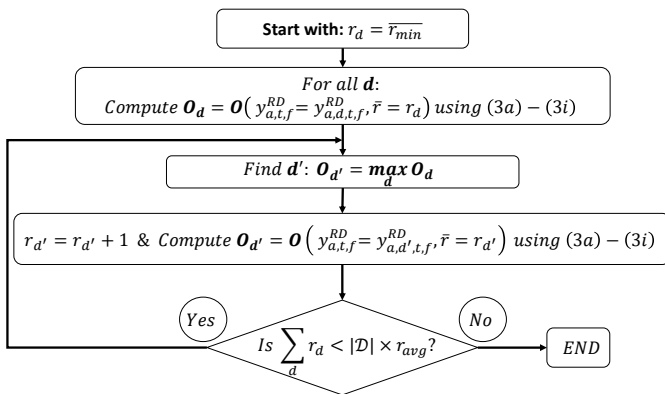


Fig. 1. Proposed optimization-based algorithm to extract RTPs from each RD

### III. CO-PLANNING MODEL FORMULATION

This section describes the proposed PWL co-planning model, both in its reference form (all days, all time points) and variations with limited RDs and RTPs.

#### A. Model Choices and Notation

This formulation is based on the notion of *areas* that consist of one or more electrical *buses* (indexed by  $b$ ). The allocation of buses to areas is represented by the matrix  $M^B$ , where  $M_{b,a}^B$  is 1 if bus  $b$  is located in area  $a$ , and 0 otherwise. Similarly, the assignment of load demand  $l$ , thermal generators  $g$ , wind power generators  $w$  and storage systems  $s$  to buses  $b$  is given by matrices  $M^L$ ,  $M^G$ ,  $M^W$  and  $M^S$ , respectively, with  $M_{b,x}^X$  indicating the assignment of resource  $x$  to bus  $b$ .

Iterations or additions over sets are commonplace in optimization problems. For notational simplicity,  $\forall x$ , where  $x$  is a relevant index, is used as shorthand for  $\forall x \in \mathcal{X}$  throughout this section: the relevant set or tuple is implicit. Indices  $b, a, l, g, w, s$  have corresponding sets  $\mathcal{B}, \mathcal{A}, \mathcal{L}, \mathcal{G}, \mathcal{W}$  and  $\mathcal{S}$ , respectively. Time points  $t$  are part of the ordered tuple  $\mathcal{T}$ , and whenever two neighboring time points (e.g.,  $t$  and  $t + 1$ ) are referenced,  $\forall t$  only refers to those  $t$  for which both are valid indices. Days  $d$  may be part of an ordered tuple (the reference case) or unordered set (RDs)  $\mathcal{D}$ . Indices  $sd$  are elements of the ordered set  $\mathcal{SD}$ .

Wind power is the only renewable resource in the present formulation, but other variable power resources such as solar PV can be added in the same way. The ESSs are implemented as long-term units that require state variables to track an annual state of charge. More generally, the presentation in this section prioritizes clarity and compactness over the maximum efficiency of resulting equations. Generic equations for investment and operation are presented in sections III-B and III-C, followed by three specializations: reference model, RD model, and RDTP (RD and RTP) model. Some redundant equations could be removed for each of the resulting models.

#### B. Investment Options

Investment options in the proposed co-planning model are new transmission lines, ESSs, and wind farms. The corresponding total cost of investment ( $CI$ ) is given by

$$CI = \sum_{nl} [CL_{nl}LL_{nl}Y_{nl}] + \sum_s [CE_sE_s^{cap} + CC_sC_s^{cap}] + \sum_w [CW_wW_w] \quad (5)$$

All investment costs are converted to equivalent annual costs using the capital recovery factor of each investment option [13].  $CL_{nl}$ ,  $LL_{nl}$  and  $Y_{nl}$  are investment cost (\$/km), length (km), and construction binary variable of new lines, respectively.  $E_s^{cap}$ ,  $C_s^{cap}$ ,  $CE_s$  and  $CC_s$  are ESS energy and power capacity variables and related investment costs, respectively.  $W_w$  and  $CW_w$  are the power capacity of the installed wind

farm and its investment cost. The investment constraints are given by

$$E_s^{cap} \leq E_s^{max}, C_s^{cap} \leq C_s^{max}, C_s^{cap} \phi_s \leq E_s^{cap} \quad \forall s \quad (6a)$$

$$0 \leq W_w \leq W_w^{max} \quad \forall w \quad (6b)$$

$$25\% \sum_l L_l \leq \sum_w W_w \quad (6c)$$

here,  $L_l$  is the peak load in each load bus. Energy and power capacity along with the energy-to-power ratio for each ESS is bounded in (6a).  $\phi_s(h)$  is the energy (MWh) to power (MW) ratio for each ESS. The capacity of wind farms, considering a renewable portfolio standard policy is limited by (6b) and (6c).

### C. Operational Model

In this subsection objective function and all related constraints for modeling operational details in reference form and variations with RDs and RTPs are presented.

1) *Operational Costs*: The cost of operation for the day  $d$  is given by

$$CO_d = \sum_t \left[ \sum_g CG_{g,d,t} + \sum_l CS_{l,d,t} \right] \quad (7)$$

where  $CG_{g,d,t}$  and  $CS_{l,d,t}$  are the cost of thermal generation and involuntary load shedding in the interval  $[t, t+1)$ , respectively. Note that we do not consider fixed operational costs in the current model, but relevant terms can readily be added. The cost of load shedding is modeled as

$$CS_{l,d,t} = \frac{VOLL}{2} (PS_{l,d,t} + PS_{l,d,t+1}) \Delta_{d,t} \quad \forall l, d, t \quad (8)$$

where  $VOLL$  is the value of lost and  $PS_{l,d,t}$  is the lost load. In this paper, thermal generation cost is modeled with a quadratic instantaneous cost curve

$$cg_g(P) = \frac{1}{2} a_g P^2 + b_g P$$

This quadratic cost function, combined with the generic PWL curve (1), results in the following cost for the interval  $[t, t+1)$  with duration  $\Delta_t$ :

$$\begin{aligned} cg_g(P_t, P_{t+1}) &= \Delta_t \int_0^1 [cg_g(P(t'))] dt' \\ &= \Delta_t cg_g\left(\frac{1}{2}(P_t + P_{t+1})\right) + \frac{a_g \Delta_t}{24} (P_t - P_{t+1})^2 \end{aligned}$$

where the result is followed by integration and collection of terms. We approximate the thermal generation costs  $CG_{g,d,t}$  by the linear lower convex envelope determined by bounding each of the two terms:

$$CG_{g,d,t} = CG_{g,d,t}^{(1)} + CG_{g,d,t}^{(2)} \quad \forall g, d, t \quad (9a)$$

$$\frac{CG_{g,d,t}^{(1)}}{\Delta_{d,t}} \geq (a_g \pi_{g,k} + b_g) \frac{PG_{g,d,t} + PG_{g,d,t+1}}{2} - \frac{1}{2} a_g \pi_{g,k}^2 \quad \forall g, k, d, t \quad (9b)$$

$$\frac{CG_{g,d,t}^{(2)}}{\Delta_{d,t}} \geq \frac{a_g \pi_{g,k}}{12} (PG_{g,d,t} - PG_{g,d,t+1}) - \frac{1}{24} a_g \pi_{g,k}^2 \quad \forall g, k, d, t \quad (9c)$$

$$\frac{CG_{g,d,t}^{(2)}}{\Delta_{d,t}} \geq -\frac{a_g \pi_{g,k}}{12} (PG_{g,d,t} - PG_{g,d,t+1}) - \frac{1}{24} a_g \pi_{g,k}^2 \quad \forall g, k, d, t \quad (9d)$$

$$\pi_{g,k} = \left( \frac{k-1}{K-1} \right) PG_g^{max} \quad \forall g, k \quad (9e)$$

where  $k \in (1, \dots, K)$  indexes the tangent lines and  $\pm \pi_{g,k}$  are the power levels for which the linear constraints (9b)-(9d) are binding. Note that  $PG_{g,d,t}$  is the thermal generator output power.

2) *Power Balance*: The local power balance equations enforce the conservation of power in every bus.

$$\begin{aligned} \sum_l M_{b,l}^L PL_{l,d,t} + \sum_s M_{b,s}^S PC_{s,d,t} = \\ \sum_g M_{b,g}^G PG_{g,d,t} + \sum_w M_{b,w}^W PW_{w,d,t} + \sum_s M_{b,s}^S PD_{s,d,t} \\ + \sum_l M_{b,l}^L PS_{l,d,t} + PN_{b,d,t} \quad \forall b, d, t \quad (10) \end{aligned}$$

here,  $PL_{l,d,t}$  and  $PW_{w,d,t}$  are bus-specific demand pattern and wind farm dispatched power.  $PC_{s,d,t}$  and  $PD_{s,d,t}$  are charging and discharging power of each ESS.  $PN_{b,d,t}$  denotes the power transported to node  $b$  by the network. The network consists of lines  $el \in \mathcal{EL}$  that are present throughout and lines  $nl \in \mathcal{NL}$  that are optionally constructed, the presence of which is indicated by the binary variable  $Y_{nl}$ . The location and orientation of lines are determined by the *directed* bus-line incidence matrices  $A_{b,el}^{EL}, A_{b,nl}^{NL}$  and the oriented power flows are given by the variables  $FE_{el,d,t}, FN_{nl,d,t}$ . Then, the flows are determined by the DC power flow equations given by

$$\begin{aligned} PN_{b,d,t} = \sum_{el} A_{b,el}^{EL} FE_{el,d,t} \\ + \sum_{nl} A_{b,nl}^{NL} FN_{nl,d,t} \quad \forall b, d, t \quad (11a) \end{aligned}$$

$$FE_{el,d,t} = B_{el} \sum_b A_{b,el}^{EL} \theta_{b,d,t} \quad \forall el, d, t \quad (11b)$$

$$-FE_{el}^{max} \leq FE_{el,d,t} \leq FE_{el}^{max} \quad \forall el, d, t \quad (11c)$$

$$\begin{aligned} -(1 - Y_{nl}) M_{nl} \leq FN_{nl,d,t} - B_{nl} \sum_b A_{b,nl}^{NL} \theta_{b,d,t} \\ \leq (1 - Y_{nl}) M_{nl} \quad \forall nl, d, t \quad (11d) \end{aligned}$$

$$-FN_{nl}^{max} Y_{nl} \leq FN_{nl,d,t} \leq FN_{nl}^{max} Y_{nl} \quad \forall nl, d, t \quad (11e)$$

where  $\theta_{b,d,t}$  is the voltage phase angle,  $M_{nl}$  is big-M for each  $nl$ , and  $B_{el(nl)}$  is the line susceptance.

3) *Generation Dispatch and Curtailment*: Thermal generators are constrained by the power and ramp constraints

$$0 \leq PR_{g,d,t} \quad \forall g, d, t \quad (12a)$$

$$PR_{g,d,t} \leq PG_{g,d,t} \leq PG_g^{max} - PR_{g,d,t} \quad \forall g, d, t \quad (12b)$$

$$\left| \frac{PG_{g,d,t+1} - PG_{g,d,t}}{\Delta_{d,t}} \right| + \frac{PR_{g,d,t}}{\tau} \leq R_g^{max} \quad \forall g, d, t \quad (12c)$$

$$\left| \frac{PG_{g,d,t+1} - PG_{g,d,t}}{\Delta_{d,t}} \right| + \frac{PR_{g,d,t+1}}{\tau} \leq R_g^{max} \quad \forall g, d, t \quad (12d)$$

where  $PR_{g,d,t}$  is the flexible spinning reserve requirement (identical up and down) allocated to the generator  $g$  and  $\tau < \Delta_{d,t}$  is its delivery time, and  $R_g^{max}$  is the maximum

ramp rate (also up and down). Eqs. (12c)-(12d) reflect the requirement to deliver the reserve contracted at the start and end of the time period, respectively<sup>1</sup>. The total ramp requirement is determined by [1]

$$\sum_g PR_{g,d,t} = 3\% \sum_l PL_{l,d,t} + 5\% \sum_w PW_{w,d,t}, \forall d, t \quad (13)$$

The bus-specific demand patterns in each area and load shedding possibility are given by

$$PL_{l,d,t} = \sum_{a,b} M_{b,a}^B M_{b,l}^L FL_{a,d,t} L_l \quad \forall l, d, t \quad (14a)$$

$$PS_{l,d,t} \leq 50\% PL_{l,d,t} \quad \forall l, d, t \quad (14b)$$

here,  $FL_{a,d,t}$  is instant load representative factors in each area. In constraint (14b) instantaneous load shedding is bounded by a percentage of demand  $PL_{l,d,t}$ . The bus-specific dispatched wind power in each area is determined by

$$PW_{w,d,t} = \sum_{a,b} [M_{b,a}^B M_{b,w}^W FW_{a,d,t} W_w] - PX_{w,d,t} \quad \forall w, d, t \quad (15a)$$

$$0 \leq PX_{w,d,t} \leq \sum_{a,b} M_{b,a}^B M_{b,w}^W FW_{a,d,t} W_w \quad \forall w, d, t \quad (15b)$$

where  $FW_{a,d,t}$  and  $PX_{w,d,t}$  represent instant wind representative factors in each area and wind curtailment.

4) *Energy Storage System*: Each ESS has grid-side power limit of  $C_s^{cap}$ , efficiencies  $\eta_s^C$  (charging) and  $\eta_s^D$  (discharging) and stored energy limit of  $E_s^{cap}$ . The evolution of stored energy levels is given by

$$E_{s,d,t+1} = E_{s,d,t} + \Delta_{d,t} \eta_s^C \frac{PC_{s,d,t} + PC_{s,d,t+1}}{2} - \frac{\Delta_{d,t}}{\eta_s^D} \frac{PD_{s,d,t} + PD_{s,d,t+1}}{2} \quad \forall s, d, t \quad (16)$$

The charge and discharge power of ESS are constrained by

$$0 \leq \eta_s^C PC_{s,d,t} \leq C_s^{cap} \quad \forall s, d, t \quad (17a)$$

$$0 \leq \frac{1}{\eta_s^D} PD_{s,d,t} \leq C_s^{cap} \quad \forall s, d, t \quad (17b)$$

$$0 = PC_{s,d,t} \times PD_{s,d,t} \quad \forall s, d, t \quad (17c)$$

Here, the complementarity constraint (17c) ensures that ESS can only charge *or* discharge at  $t$ . We point out that in the PWL model, the complementarity constraint is only strictly enforced at time points  $t$ , but it is possible for the charging and discharging power to have (implied) nonzero values in between these time points. As this leads to a slightly conservative result (due to additional losses), we omit further constraints that would introduce additional nonlinearities. (17c) is implemented using binary variables as:

$$\eta_s^C PC_{s,d,t} \leq U_{s,d,t} C_s^{max} \quad \forall s, d, t \quad (18a)$$

$$\frac{1}{\eta_s^D} PD_{s,d,t} \leq (1 - U_{s,d,t}) C_s^{max} \quad \forall s, d, t \quad (18b)$$

<sup>1</sup>We note that (12c) differs from the power-based start-of-hour ramping formulation proposed in [12].

## D. Reference Model

In the reference model, stored energy levels are limited by

$$0 \leq E_{s,d,t} \leq E_s^{cap} \quad \forall s, d, t \quad (19)$$

Moreover, the  $d$  indices refer to sequential days in the data set. Therefore, we impose the following boundary constraints on *all* variables  $X_{y,d,t}$  that have indices  $d$ , and  $t$  (and an  $X$ -specific index  $y$ ), to impose continuity between days and between the first and last day:

$$X_{y,d+1,0} = X_{y,d,24} \quad \forall y, d \quad (20a)$$

$$X_{y,1,0} = X_{y,|\mathcal{D}|,24} \quad \forall y \quad (20b)$$

The objective function is to minimize the *total planning cost*, consisting of the investment and operational parts.

**Reference Model:**

$$TPC^{\text{ref}} = \min [CI + \sum_{d \in \mathcal{D}} CO_d] \quad (21)$$

$$s.t.: \quad \text{constraints (5)-(20),}$$

$$\mathcal{D} = (\text{all days, sequentially}),$$

$$\mathcal{T} = (0, \dots, 24),$$

$$\Delta_{d,t} = 1 \text{hr} \quad \forall d, t$$

## E. Representative Day Model

When using RDs, the long-term ESS model must be adapted accordingly. We follow the ERD (Enhanced Representative Days) methodology [6] with the extension to SLDs proposed in [1]. In this representation, the variables  $E_{s,d,t}$  no longer represent the absolute state of the charge of the ESS, but its *change* relative to the start of RD  $d$ . We define its total change in energy  $E_{s,d}^{\text{tot}}$  and its minimum/maximum excursion during the day as follows:

$$E_{s,d}^{\text{tot}} = E_{s,d,24}, \quad E_{s,d,0} = 0 \quad \forall s, d \quad (22a)$$

$$E_{s,d}^{\text{low}} \leq E_{s,d,t} \leq E_{s,d}^{\text{high}} \quad \forall s, d, t \quad (22b)$$

In addition, each ESS is assigned a variable  $LE_{s,d}$  to track its state of charge at the beginning (and end) of each SLD. The association of RDs and SLDs is given by the matrix  $D_{sd,d}$ , with element values  $\{0, 1\}$  and the number of RD repetitions within an SLD is  $n_{sd}^B$ . The long-term stored energy is tracked for every SLD, using the following equations:

$$LE_{s,sd+1} = LE_{s,sd} + n_{sd}^B \sum_d D_{sd,d} E_{s,d}^{\text{tot}} \quad \forall s, sd \quad (23a)$$

$$LE_{s,1} = LE_{s,|\mathcal{SD}|} + n_{|\mathcal{SD}|}^B \sum_d D_{|\mathcal{SD}|,d} E_{s,d}^{\text{tot}} \quad \forall s \quad (23b)$$

$$0 \leq LE_{s,sd} + \sum_d D_{sd,d} E_{s,d}^{\text{low}} \quad \forall s, sd \quad (23c)$$

$$0 \leq LE_{s,sd} + \sum_d D_{sd,d} [(n_{sd}^B - 1) E_{s,d}^{\text{tot}} + E_{s,d}^{\text{low}}] \quad \forall s, sd \quad (23d)$$

$$E_s^{cap} \geq LE_{s,sd} + \sum_d D_{sd,d} E_{s,d}^{\text{high}} \quad \forall s, sd \quad (23e)$$

$$E_s^{cap} \geq LE_{s,sd} + \sum_d D_{sd,d} [E_{i,d}^{\text{high}} + (n_{sd}^B - 1) E_{s,d}^{\text{tot}}] \quad \forall s, sd \quad (23f)$$

Combining all the above constraints results in the following RD model in PWL representation.

### Representative Day Model (RD):

$$\begin{aligned}
 TPC^{\text{RD}} &= \min [CI + \sum_{d \in \mathcal{D}} \omega_d CO_d] & (24) \\
 s.t : & \text{ constraints (5)-(18), (22)-(23),} \\
 & \mathcal{SD} = (1, \dots, \# \text{ of SLDs}), \\
 & \mathcal{D} = \{\text{RDs}\}, \\
 & \omega_d = \{\text{weight of each RD}\}, \quad \forall d \\
 & \mathcal{T} = (0, \dots, 24), \\
 & \Delta_{d,t} = 1\text{hr} \quad \forall d, t
 \end{aligned}$$

### F. Representative Day & Time Point Model

Finally, the RDTP model can be constructed from the previously presented constraints, using a set of RDs and unequal time points computed by (3), either using equal or unequal allocation to of RTPs to RDs. The main difference with the RD model is that the set of time indices  $\mathcal{T}_d$  now depends on the RD  $d$ . This also necessitates updating (22a) to

$$E_{s,d}^{\text{tot}} = E_{s,d,\max(\mathcal{T}_d)} \quad \forall s, d \quad (25)$$

The final model is then given by:

### Representative Day & Time Point Model (RDTP):

$$\begin{aligned}
 TPC^{\text{RDTP}} &= \min [CI + \sum_{d \in \mathcal{D}} \omega_d CO_d] & (26) \\
 s.t : & \text{ constraints (5)-(18), (22b), (23), (25),} \\
 & \mathcal{SD} = (1, \dots, \# \text{ of SLDs}), \\
 & \mathcal{D} = \{\text{RDs}\}, \\
 & \omega_d = \{\text{weight of each RD}\}, \quad \forall d \\
 & \mathcal{T}_d = (0, \dots, r_d - 1), \quad \forall d \\
 & \Delta_{d,t} = \text{variable, according to equation (4)} \quad \forall d, t
 \end{aligned}$$

## IV. PERFORMANCE ANALYSIS

### A. Study System

The effectiveness of the proposed co-planning model and representative period extraction methods was evaluated using the IEEE RTS 24-bus test system [14]. We considered seven areas with distinct load and RES generation patterns based on the Netherlands (buses 9, 11, 15, 16, 26), Belgium (buses 17, 18, 25), France (buses 1, 3, 4, 24), Germany (buses 6, 10, 12, 13), Denmark (buses 14, 19, 20), Sweden (buses 21, 22, 23), and Switzerland (buses 2, 5, 7, 8). Note that buses 25 and 26 with candidate wind farms are assumed to be expansion buses. All the data for this test system, including system topology, cost of existing and new candidate options, required parameters, along with data source references, are available in [15]. The one-year load and renewable energy generation data for each area are based on one year of data from 2019. The CPLEX solver in the GAMS environment [16] was employed to solve the proposed MILP co-planning and optimization-based RTP extraction problems. Additionally, the multi-area clustering-based RD extraction algorithm was implemented in Matlab [17], running on a PC with an Intel Xeon W-2223 CPU 3.60 GHz and 16 GB of RAM.

TABLE I  
RESULTS FOR THE REFERENCE MODEL (REF-PWL)

| Cost:                      | TCO*                             | CI           | TPC           |                     |
|----------------------------|----------------------------------|--------------|---------------|---------------------|
| Value ( $\times 10^6$ \$): | 2021.4                           | 776.4        | 2797.8        |                     |
| Planning Option:           | Transmission Line                | ESS (bus)    |               | Wind Farm (bus)     |
|                            |                                  | Short-term   | Long-term     |                     |
| Location:                  | (7-8), (14-16), (19-25), (16-26) | 2, 3, 16, 25 | 8, 11, 19, 23 | 3, 5, 6, 14, 25, 26 |
| Load Shedding (GWh)        | 4.92                             |              |               |                     |
| CPU Time (Sec):            | 242, 825                         |              |               |                     |

\*: Total Cost of Operation

### B. Numerical Results

The planning problem was solved using different model formulations. First, the reference model **Ref-PWL** (21) was solved using all hours and days, with results shown in Table I. Four new transmission lines, four short-term ESS, and four long-term ESS, along with six wind farms were scheduled for installation and the required CPU time was  $\approx 67$  hours, highlighting the importance of problem reduction.

To facilitate a comparison between PWL and PWC models, an equivalent PWC reference model was constructed, by reformulating equations (8), (9a)-(9d), (16) and (20) to use 24 constant power values. The investment decisions for the reference PWC model were identical, and the other costs were comparable with the PWL model (within the optimality gap). However, the advantages of the PWL approach become evident when approximating the model in reduced space cases with RDs and RTPs.

Models with representative days and/or time periods were compared to their respective reference models on the basis of investment cost, operation cost, and total cost. The error metric used to analyze the effectiveness of the proposed method is

$$\text{Error} = \frac{f_c^*(\hat{v}) - f_c^*(v^*)}{f_c^*(v^*)}. \quad (27)$$

Here  $v^*$  represents the decision variables for the reference case, while  $\hat{v}$  denotes the fixed decision variables obtained from the reduced space models.  $f_c^*$  signifies the desired cost type, assessed across all days and hours (non-reduced).

In cases **RD-noEx** and **RD**, the PWL RD model (24) was evaluated with 21 extracted RDs, ignoring and considering extreme value days, respectively. The number of resulting SLDs was 271 for **RD-noEx**, and 267 for **RD**. The results in Table II illustrate the significance of capturing extreme value days. The error in total planning cost reduced dramatically from nearly 3% to 0.23%, resulting from reductions in investment

TABLE II  
ERROR AND CPU TIME COMPARISON BETWEEN CASES RD-NOEX & RD

| Case           | Error (%) |            |       | CPU Time (sec): |
|----------------|-----------|------------|-------|-----------------|
|                | Operation | Investment | Total |                 |
| <b>RD-noEx</b> | 6.214     | -5.312     | 2.908 | 414             |
| <b>RD</b>      | -0.041    | 0.945      | 0.234 | 637             |

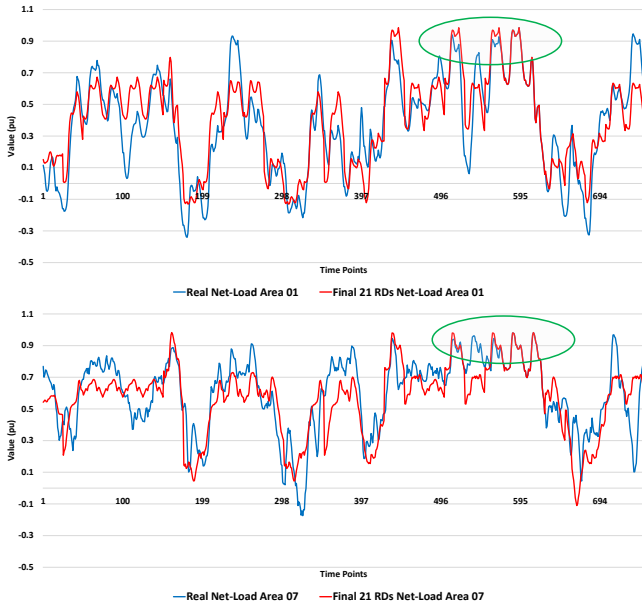


Fig. 2. Comparison of original net load and reconstructed net load based on RDs, for two areas (top and bottom), for one month with hourly resolution. The ability to capture net load peaks using extreme day preservation is highlighted

and operation cost errors. Fig. 2 highlights the ability of the extreme-day sensitive RD selection process to capture net load peaks. The use of RDs reduced CPU time by up to 99.7%. For subsequent results, the extreme day selection was used.

The efficacy of the *RDTP model* (26) was investigated next. The benefit of an adaptive allocation of RTPs across RDs can be illustrated in two other ways. First, as presented in Table III, after executing the proposed optimization-based method (3)-(4) to extract 10 RTPs within each of the 21 extracted RDs, for a total of 210 RTPs, the adaptive selection of numbers of RTPs resulted in an average total absolute error (i.e., (3a)) of 1.578 pu, compared to a value of 1.602 pu for the case where each RD was allocated exactly 10 RTPs. Additionally, Fig. 3 provides an illustrative example, comparing both methods

TABLE III  
EQUAL AND ADAPTIVE NUMBER OF RTPs COMPARISON WITHIN EACH RD

| RD:                     | #of Extracted RTP |           | RD:               | #of Extracted RTP |           |
|-------------------------|-------------------|-----------|-------------------|-------------------|-----------|
|                         | Equal:            | Adaptive: |                   | Equal:            | Adaptive: |
| 1                       | 10                | 09        | 12                | 10                | 12        |
| 2                       | 10                | 10        | 13                | 10                | 11        |
| 3                       | 10                | 10        | 14                | 10                | 10        |
| 4                       | 10                | 09        | 15                | 10                | 11        |
| 5                       | 10                | 10        | 16                | 10                | 09        |
| 6                       | 10                | 10        | 17                | 10                | 08        |
| 7                       | 10                | 11        | 18                | 10                | 09        |
| 8                       | 10                | 11        | 19                | 10                | 09        |
| 9                       | 10                | 11        | 20                | 10                | 09        |
| 10                      | 10                | 13        | 21                | 10                | 09        |
| 11                      | 10                | 10        | TNRT <sup>*</sup> | 210               | 210       |
| ATAE <sup>**</sup> (pu) |                   |           |                   | 1.602             | 1.578     |

\*: Total Number of RTPs, \*\*: Average of Total Absolute Error

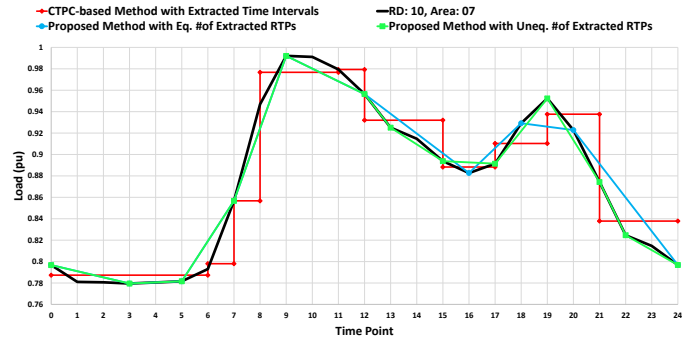


Fig. 3. Comparison of extraction of equal and unequal number of RTPs for PWL models, alongside the CTPC method for PWC models

of extracting equal and unequal numbers of RTPs with the CTPC method, for load data of area 07 in RD 10. The adaptive method assigned 13 RTPs to this relatively variable day, allowing it to capture more details. Moreover, both PWL-based methods are able to better approximate ramps, even when larger time steps occur between RTPs.

The first case, **RD-CTPC**, is based on a PWC formulation. 10 representative time intervals were extracted from each of the 21 extracted RDs using the CTPC method [7]. Planning errors were calculated using (27) by comparison with the PWC reference model and the obtained results are presented in Table IV. A significant reduction in CPU time is observed due to the use of RDs.

In cases **RDTP-eq** and **RDTP-ad**, the proposed optimization-based method (3a)-(4) was utilized to extract 10 RTPs, with equal and unequal RTPs within each of the 21 RDs, respectively. The results were compared to Ref-PWL in Table IV. Both PWL-based methods outperform the PWC-based approach with CTPC, and the adaptive allocation of RTPs across RDs yielded a slight further improvement. Compared to the RD methods, a speedup of approximately 6 times was obtained, at a very minor reduction in accuracy (comparing **RD** with **RDTP-ad**). Compared to the reference case **Ref-PWL**, the TPC and CI in case **RDTP-ad** increased by 0.348% and 1.449%, respectively, and the TCO decreased by 0.075%, as also presented in Table IV. Similar to the **RDTP-eq** method, the CPU time was significantly reduced for both RDTP methods. Jointly, the reduced error and large computational savings confirm the effectiveness of the proposed **RDTP-ad** method in balancing complexity and

TABLE IV  
ERROR AND CPU TIME COMPARISON BETWEEN CASES RD-CTPC, RDTP-EQ & RDTP-AD

| Case    | Error (%) |            |       | CPU Time (sec): |
|---------|-----------|------------|-------|-----------------|
|         | Operation | Investment | Total |                 |
| RD-CTPC | -0.210    | 2.571      | 0.562 | 84              |
| RDTP-eq | -0.089    | 1.712      | 0.411 | 97              |
| RDTP-ad | -0.075    | 1.449      | 0.348 | 95              |

TABLE V  
RESULTS FOR CASE RDTP-AD

| Cost:                      | TCO                              | CI           |               | TPC                 |
|----------------------------|----------------------------------|--------------|---------------|---------------------|
| Value ( $\times 10^6$ \$): | 2019.88                          | 787.65       |               | 2807.53             |
| Planning Option:           | Transmission Line                | ESS (bus)    |               | Wind Farm (bus)     |
|                            |                                  | Short-term   | Long-term     |                     |
| Location:                  | (7-8), (14-16), (19-25), (16-26) | 2, 3, 16, 25 | 8, 11, 19, 23 | 3, 5, 6, 14, 25, 26 |
| Load Shedding (GWh)        | 4.71                             |              |               |                     |
| CPU Time (Sec):            | 95                               |              |               |                     |

accuracy.

Table V shows the investment decision variables, along with TCO, CI, and TPC for case **RDTP-ad**. The constructed transmission lines, as well as the installed ESS and wind farm locations, were the same for both cases **Ref-PWL** and **RDTP-ad**. The reason for the different costs is attributed to the installed capacities of the ESS and wind farms. In case **RDTP-ad**, more capacities were installed to cover extreme values in the data, which led to a higher CI and lower TCO. Consequently, load shedding was also slightly less in case **RDTP-ad**.

## V. CONCLUSION

This paper proposed a hybrid multi-area piecewise linear (PWL) adapted method for capturing interday and intraday chronology, considering extreme values. This was achieved by extracting representative days (RDs) and time points (RTPs) within each RD, addressing a complex co-planning problem involving transmission lines, energy storage systems, and wind farms. To enhance the representation of intraday chronology, an optimization-based RTP extraction approach was introduced that adaptively extracts unequal numbers of RTPs within each RD. The effectiveness of the proposed model with PWL OPF formulations was evaluated using six different cases. The importance of capturing extreme values was demonstrated. Moreover, the ability of the adaptive method to assign more RTPs to RDs with higher hourly variation facilitates the accurate representation of intraday chronology. In the case study, the combination of these approaches resulted in a time saving of  $> 99.9\%$ , accompanied by a total planning cost error of only 0.348%.

Future studies might investigate dealing with a decade of input data for larger areas and additional renewable energy sources, such as solar, can be explored to further validate and expand our findings. Moreover, for optimal deployment of the method across a range of practical scenarios, it will be important to establish rules of thumb on how many RTPs to identify across which number of RDs, for a given computational budget.

## REFERENCES

[1] M. Moradi-Sepahvand and S. H. Tindemans, "Capturing chronology and extreme values of representative days for planning of transmission lines and long-term energy storage systems," in *2023 IEEE Belgrade PowerTech*, 2023, pp. 1–6.

[2] K. Poncelet, H. Höschle, E. Delarue, A. Virag, and W. D'haeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1936–1948, 2016.

[3] S. Gao, B. Hu, K. Xie, T. Niu, C. Li, and J. Yan, "Spectral clustering based demand-oriented representative days selection method for power system expansion planning," *International Journal of Electrical Power & Energy Systems*, vol. 125, p. 106560, 2021.

[4] C. Li, A. J. Conejo, J. D. Siirola, and I. E. Grossmann, "On representative day selection for capacity expansion planning of power systems under extreme operating conditions," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107697, 2022.

[5] A. Orgaz, A. Bello, and J. Reneses, "Modeling storage systems in electricity markets with high shares of renewable generation: A daily clustering approach," *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107706, 2022.

[6] S. Gonzato, K. Bruninx, and E. Delarue, "Long term storage in generation expansion planning models with a reduced temporal scope," *Applied Energy*, vol. 298, p. 117168, 2021.

[7] S. Pineda and J. M. Morales, "Chronological time-period clustering for optimal capacity expansion planning with storage," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7162–7170, 2018.

[8] A. Garcia-Cerezo, R. Garcia-Bertrand, and L. Baringo, "Priority chronological time-period clustering for generation and transmission expansion planning problems with long-term dynamics," *IEEE Transactions on Power Systems*, 2022.

[9] H. Teichgraber and A. R. Brandt, "Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 157, p. 111984, 2022.

[10] E. Nycander, G. Morales-España, and L. Söder, "Power-based modelling of renewable variability in dispatch models with clustered time periods," *Renewable energy*, vol. 186, pp. 944–956, 2022.

[11] D. A. Tejada-Arango, G. Morales-España, S. Wogrin, and E. Centeno, "Power-based generation expansion planning for flexibility requirements," *IEEE Transactions on Power Systems*, vol. 35, no. 3, 2019.

[12] E. Nycander, G. Morales-España, and L. Söder, "Security constrained unit commitment with continuous time-varying reserves," *Electric Power Systems Research*, vol. 199, p. 107276, 2021.

[13] M. Moradi-Sepahvand and T. Amraee, "Secure expansion of energy storage and transmission lines considering bundling option under renewable penetration," *Applied Energy*, vol. 347, p. 121414, 2023.

[14] C. Grigg, P. Wong, P. Albrecht, R. Allan, M. Bhavaraju, R. Billinton, Q. Chen, C. Fong, S. Haddad, S. Kuruganty, W. Li, R. Mukerji, D. Patton, N. Rau, D. Reppen, A. Schneider, M. Shahidehpour, and C. Singh, "The IEEE reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1010–1020, 1999.

[15] [Online]. Available: <https://github.com/Mojtabamoradi/Modified-RTS-24-bus>.

[16] "The general algebraic modeling system." [Online]. Available: <https://www.gams.com/>.

[17] "The math works, inc. matlab. version r2020b." [Online]. Available: <https://www.mathworks.com/>.