Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Performance measures and availability of European power plants

Eric Jahnke
eric.jahnke@kit.edu

Karlsruhe Institute of Technology

Armin Ardone
Karlsruhe Institute of Technology

data-descriptor

Keywords:

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9138908/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Reliable representation of power plant availability is essential for energy system analysis, electricity market modelling and security-of-supply studies, yet openly accessible datasets are scarce and heterogeneous. This data set presents a harmonised set of performance measures for thermal and hydropower units in the ENTSO-E area derived from monthly unavailability reports on the ENTSO-E Transparency Platform. We reconstruct unit-level availability time series from event-based outage records, apply transparent filtering, clustering and interpolation procedures, and enhance the data with technology and commissioning information from merged open power plant databases. From these processed time series we compute hourly measures and a broad range of time-based performance indicators, including availability factors, outage factors and rates, failure and repair rates, and statistics on partial derations, in both capacity-weighted and unweighted form. The resulting indicators are provided by country, season, plant type and technology, and are validated against historical generation records, selected adequacy studies as well as other journal publications. The accompanying open-source Python workflow enables full reproducibility and direct reuse in data processing or modelling. In addition, two example algorithms are included that can be used to simulate outage series for individual generation units or aggregated power plant fleets.

Background & Summary

Availability data of power units is required for realistic models in the energy industry or electrical engineering for various applications e.g., to map supply shortages in electricity market models, to provide the basis for investment decisions based on security of supply in energy system models or to analyse contingency cases in power flow optimizations. On detailed technology base and for specific countries, licensed data is available[1]. However, publicly available data sources are scarce. This data set provides aggregated statistics and performance measures for European electricity generation technologies and fuel types. Information about the performance of units is based on the monthly unavailability reports of the ENTSO-E transparency platform[2]. Methodologically, the data preparation is comparable to other recent publications in this field[3,4]. A new feature is a reproducible Python pipeline to generate regular time series per unit, more optional aggregations as well as validation against different studies. The performance measures comply with the current IEEE-762 standard[5] and other KPIs used in practice[6]. The data set can thus be used directly for applications such as probabilistic outage sampling, the simulation of time series in different granularity or as constraints for the aforementioned optimization models.

Methods

Step 1 - Data source

We use monthly CSV file exports of the ENTSO-E transparency platform ("*UnavailabilityOfGenerationUnits_15.1.A_B*"). Additional columns of clear names of the ENTSO-E identification codes of bidding zones or control areas were created using mapping tables taken from the Python API for the ENTSO-E transparency data download[7]. The data only contains entries from generation units with a nominal capacity of at least 100 MW and outage reports with a change in capacity of at least 100 MW. Outages of consumption units or transmission lines are not included, as there are other data sources for this on the transparency platform[8]. In November 2025 when this study was written, the ENTSO-E updated its data access from the FTP server and, in the process, also changed the table formats of some data sources[9]. As the code and statistics were created when the unavailability forms were still in change and the final format unclear, an overview and example of the old and currently proposed new table format is provided in **Table 1**. The time series for the outage notification is taken from the "StartOutage(UTC)" and "EndOutage(UTC)" columns. The columns "StartTimeSeries(UTC)" and "EndTimeSeries(UTC)" are ignored, as no further information is available for them.

Old column name	New column name	Data type	Description	Example
StartOutage	StartOutage(UTC)	*Timestamp*	**Start time of an outage event**	2025-11-12 13:30:00.000
EndOutage	EndOutage(UTC)	*Timestamp*	**End time of an outage event**	2025-11-12 23:30:00.000
StartTS	StartTimeSeries(UTC)	*Timestamp*	**Start of a change of the avail. cap. during an outage**	2025-11-12 15:30:00.000
EndTS	EndTimeSeries(UTC)	*Timestamp*	**End of a change of the avail. cap. during an outage**	2025-11-12 20:30:00.000
TimeZone	TimeZone	*String*	**Local time zone (e.g., "CET")**	CET
MRID	InstanceCode	*String*	**Unique identifier of outage event**	123XYZ
Status	Status	*String*	**Current status**	Cancelled
Type	Type	*String*	**"Planned" or "Forced" outage event**	Forced
AreaCode	AreaCode	*String*	**Unique identifier of market area**	10YGB———-A
AreaTypeCode	AreaTypeCode	*String*	**Bidding zone or control area ("BZN" or "CTA")**	CTA
AreaName	AreaDisplayName	*String*	**Name of BZN or CTA**	UK(National Grid) CTA
MapCode	AreaMapCode	*String*	**Unique identifier of BZN or CTA**	GB
PowerResourceEIC	AssetCode	*String*	**Unique identifier of the unit**	48W0000ABC-15A
UnitName	AssetName	*String*	**Name of the unit**	GBX-2
ProductionType	ProductionType	*String*	**Fuel or tech. type (e.g., "Nuclear")**	Fossil Oil
InstalledCapacity	InstalledCapacity	*Float*	**Installed capacity**	200
AvailableCapacity	AvailableCapacity[MW]	*Float*	**Available capacity during the outage event**	100
Version	Version	*Integer*	**Update version of the MRID**	2
OldVersion	OldVersion	*Boolean*	**Indicator if the version is still the most current**	FALSE
Reason	Reason	*String*	**Reason for the outage (e.g., "Failure")**	Failure
UpdateTime	UpdateTime(UTC)	*Timestamp*	**Update time of the outage event**	2025-11-12 12:30:00.000

Table 1: Overview about raw monthly CSV file for unavailability of generation units from the old as well as newly updated format[2,8,10].

Step 2 - Data cleaning

Not all single entries are valid anymore. If maintenance is planned in the medium term due to a technical problem that has occurred and this can be resolved in the short term, the original notification for future maintenance work is cancelled. As a result, the raw data set is filtered exclusively for reports that are labelled as "Active" and not "Withdrawn" or "Cancelled". Regardless of their type, outage notifications have an identification code ("MRID") for clear traceability with, for example, the additional ENTSO-E table for the reasons of the outage. However, as outages for planned maintenance work have to be postponed or ad-hoc messages for unplanned outages have to be readjusted at short notice due to technical problems, there are often duplicates for a specific identification code. Consequently, all entries that show an old version were filtered out. As there were still duplicates in the identifier, all reports were sorted by version and only the line with the highest version number was taken over. In this respect, the clean-up approach of Gjorgiev et al.[3] differs, as the entries with duplicate identifiers were not removed after repeated filtering, but were treated as different reports with more accidentally similar identifiers. In our approach, if there were still duplicates of the identifiers, the entry with the most recent update time was used[1]. All duplicated entries are collected during processing and can be written out as a CSV file. Some countries are missing entirely in the bidding zones (e.g., Germany). Since Germany is also divided into four balancing zone, there are four entries instead of one aggregated. For this reason, we do not yet filter specifically by bidding zone or control areas during data preparation, but instead specify the countries to be filtered by before calculating the performance indicators. There are mainly duplicates here because we did not filter bidding zones or control areas. An overview of the raw and cleaned dataset of reports used for this study is given in **Table 2**:

	N
Raw reports	12,042,695
Cancelled	967,496
Withdrawn	301,275
Old version	9,466,610
Active current version	1,307,314
Duplicated rows	287,742
Duplicated MRIDs	69,975
Active unique reports	949,597
Active filtered reports	**947,061**[2]
Duplicated EIC codes	6
Long-ranging single outage clusters	315
Reports used for performance measures	**946,740**

Table 2: Overview of raw outage reports before and after cleaning and pre-processing

Step 3 - Data pre-processing

Bridging temporal gaps

Sometimes there is a temporal gap of only a few hours between two outage reports, where it can be assumed that the generation unit was not running in normal operation and this gap evolved probably from a reporting error. For this reason, a time interval of τ hours can be defined in which two failure reports are linked to each other (**Figure 1**). The possibility of defining a bridge parameter is optionally available in the accompanying code, whereby the value of the available capacity as well as the type and reason for the earlier outage are extended until the later outage. Optionally, parameters can also be set to link only short time deltas between reports of the same type and/or reason. A delta of eight hours was assumed for the calculation of the statistics for this data set, but even more hours would have been possible according to sensitivity analysis in other studies[3].

Clustering and Labelling

Outage events for a generation unit often overlap in time. The problem here is that not only different available capacities are reported for a specific point in time, but the type and reason for the outage can also differ. In addition, the reports have different creation or update times, which makes it difficult to clearly identify which is the current report in the event of a temporal overlap. The possible reasons for the outage are divided by us into just two categories: “Maintenance” and “Other”. A more specific differentiation would not bring more benefit to the statistical analysis. Regarding the type, all reports distinguish between planned and forced outages. Officially, the report for a planned unavailability must be published to the transparency platform maximum one hour after the official confirmation by the TSO. Updates to this report (a newer version) must be published at the latest an hour after the information is known. Information on unplanned outages must be published no later than one hour after the change in available capacity[8]. In principle, this is in line with other institutions such as NERC, whose declarations between a long-term planned outage, planned maintenance or unplanned maintenance depend on the announcement time before the event and may have to be adjusted afterwards[11–13]. However, for the ENTSO-E unavailability data the difference between notifications for forced and planned is often unclear, as planned outages are only partially published with a long lead time[4]. In some cases, reports of planned maintenances don’t even exist, because every maintenance is reported as ‘forced’ (a more detailed analysis can be found in the **supplementary materials**). In reality, the correct classification of outage types as forced or planned, as well as the separation of overlapping reports, depends on the type of power plant and specific background knowledge about the reasons for the outages in terms of expected repair times[11]. The labelling is therefore subject to a level of uncertainty that might bias the statistical analysis.

Previous works used different heuristics and methods to cluster and label the reports. Schmitz et al.[14] and Finck[15] aggregate the (un-)availabilities of the reports regardless if it was planned or forced. Gils et al.[16] or Deakin et al.[17] differentiate only between planned and forced, but they ignore the outage reasons and do not explain their methodology for pre-processing the data. Bassini[4] characterises the type in the event of a temporal overlap of 3 notifications based on the majority vote. The procedure for an even number of overlapping events such as two, four or more is not described. We differentiate between the clustering of the overlapping outage reports with regard to the value of the available capacity, the label for the outage type, and the reason. For each time step with overlapping reports, we assign the value of the available capacity by the report with the most recent update timestamp. For the label of the outage type and reason, we assign the values for the earliest outage. For reports with identical outage starts, we again assign the label with the most recent update time. However, the same logic can be implemented for the labels as for the capacity using

parameters in the code, so that the report with the latest update timestamp is always decisive. The separation of a non-availability event into separate events is implemented when the state transitions from a planned deration or outage to a forced deration or outage[3]. An example of three overlapping outage reports for a specific day with different update times and types is shown in **Figure 2**, visualising the decisions of clustering and labelling the reports. Other cases can be found in the **supplementary materials**.

States

A generating unit can be either fully available, derated or fully unavailable. It is fully available if it can potentially produce electricity at its rated power. If a unit is in outage, it cannot generate any electricity at all. Derated means that a unit cannot operate at full capacity, but is not off the grid. Consequently, the possible states could be defined as planned outage, planned derated, unplanned outage, unplanned derated and available[3]. Since we record the outage type separately, we only define the primary states as “out”, “derated” and “available”.

Final cleaning

Since the outage data for bidding zones and control zones in countries with multiple zones (i.e., Norway and Italy) often contain overlapping entries and also show changes in the affiliation of individual power plants over time, the performance measures were only calculated for each country and duplicate entries of the EIC codes were removed beforehand. These duplicates are exported to a separate CSV file for overview purposes. Similarly, individual outages of the same type (“planned”) lasting longer than 365 days were assumed to be erroneous entries and were therefore removed. However, if two or more outages of different types (a “planned” outage followed immediately by an “unplanned” outage) exceed the defined duration, they are not removed, as they essentially represent two separate clusters[3]. All removed entries in this way are recorded and exported in a CSV file.

Step 4 - Meta information for power units

On the ENTSO-E Transparency Platform, thermal power plants in particular are only listed by fuel type and general plant types, but not broken down in more detail by technology type such as open-cycle gas turbine and combined-cycle gas turbine. The year of commissioning (or retrofit) of the units is also not available, nor is the information as to whether it is a CHP plant. We add this information from freely available data sources. To do this, we use the open units table from JRC[18] and a power plant list from the PyPSA modelling framework package *powerplantmatching*[19]. As the JRC table also contains the official EIC codes of ENTSO-E for power plant locations and the individual units, we can join the power plant lists from JRC and PyPSA together with ENTSO-E using these columns. A geographical overview of plant types, technologies and installed capacities can be seen in **Figure 3**. Descriptive statistics on the meta-information of the power plants can be found in the **supplementary materials**. An example is given with **Table 3**. Since the JRC and PyPSA power plant lists do not contain the relevant technology types or commissioning data for all units in the ENTSO-E tables, the characteristics of individual technology types are unfortunately sometimes very limited and less reliable. This applies in particular to thermal power plants that use natural gas, coal, or oil as fuel. For this reason, the results also include the values for NAs in the technologies.

Column name	Data type	Description	Example
unit_eic	*string*	**Unique identifier of the gen. unit**	48W0000ABC-15A
plant_eic	*string*	**Unique identifier of the gen. plant (a plant can consist of multiple units)**	48W0000ABC
unit_name	*string*	**Name of the unit**	GBX-2
plant_name	*string*	**Name of the plant**	GBX
plant_tech	*string*	**Technology type**	CCGT
fuel_type	*string*	**Fuel type**	Fossil gas
fuel_type_code	*string*	**ENTSO-E code of the fuel type**	B04
set	*string*	**Technology group**	CHP
year_commissioned	*integer*	**Year of first operation (or retrofit)**	1995
year_commissioned	*integer*	**Year of decommissioning**	2040
unit_installed_capacity	*float*	**Installed capacity (unit)**	400 MW
plant_installed_capacity	*float*	**Installed capacity (plant)**	800 MW
country	*string*	**Country location**	GB
nuts2	*string*	**NUTS 2 code of the unit location**	GB5
lat	*Float*	**Geographical latitude**	51.50
lon	*float*	**Geographical longitude**	-0.12
eic_match_source	*string*	**Source column to merge PyPSA on JRC power plant databases**	plant_eic

Table 3: Overview about the merged power plant list to enhance the statistical analysis.

Data Records

Repository

Code and datasets are hosted online and can be downloaded from **Zenodo**. The code already contains implementations for calculating more aggregated KPIs by e.g., age, size, or CHP type of the units. The performance measures originate from the IEEE 762 standard[5] as well as frequently used statistics in practice and academic studies[6]. They primarily relate to the time-based availability of power generation units. As the publication of data of individual non-anonymised generation units of the transparency platform is not permitted, capacity-weighted aggregated performance factors were calculated. Although these do not allow the same level of detail as plant-specific data, they are still sufficient enough for simulations of power plant outages or for a rough estimate of the availability. The files are:

- ‘kpis_tech_ALL.csv’
- ‘kpis_tech_MOY.csv’
- ‘kpis_plant_ALL.csv’
- ‘kpis_plant_MOY.csv’

The four files contain average annual and monthly performance measures for ENTSO-E plant types or technology types and cover the ENTSO-E area with the following countries: Albania (AL), Austria (AT), Belgium (BE), Bosnia-Herzegovina (BA), Bulgaria (BG), Switzerland (CH), Czechia (CZ), Germany (DE), Denmark (DK), Estonia (EE), Spain (ES), France (FR), Finland (FI), Great Britain (GB)[3], Greece (GR), Croatia (HR), Hungary (HU), Ireland (IE), Italy (IT), Lithuania (LT), Latvia (LV), Luxembourg (LU), Moldova (MD), Montenegro (ME), North Macedonia (MK), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Serbia (RS), Sweden (SE), Slovenia (SI), Slovakia (SK) and Kosovo (XK). The data is calculated as annual averages and seasonally disaggregated by month, among other things, because seasonal factors such as outside temperature, cooling water temperature, etc. change the likelihood of unplanned outages and therefore also TSOs schedule their planned maintenance according to the season[20–24].

The repository additionally contains the Python and R scripts for preparing, analysing, evaluating and visualising the data, as well as an example script for simulating further time series:

- ‘availability_data.py’
- ‘availability_statistics.py’

- ‘availability_simulation.py’
- ‘powerplants_merge.py’
- ‘plot_outages_vs_gen.py’
- ‘validate_kpis.R’
- ‘descriptive_stats_jrc_pypsa.R’
- ‘plot_outage_statistics.R’

Tables

The performance measures described above contain can be found with various additional prefixes or letter combinations as shown in **Table 4**:

Abbreviation	Example
w/u	Weighted vs. Unweighted
(E)	Equivalent
F/P/M/U/S	Abbreviation for outage type
O/D	Total outage vs. derated
Actual measure	Rate, Factor, etc.

Table 4: Structure of the name of a measure.

Outage types were taken into account according to their type and reason[5]:

- **U**nplanned - all forced outage events including technical failures as well as maintenance work.
- **S**cheduled - all planned outage events combining maintenance work with other planned outages.
- **M**aintenance - calculated accordingly from events with the reason ‘maintenance’, whereby the type can be planned or forced.
- **P**lanned - calculated from reports containing planned outages where the reason is not maintenance.
- **F**orced - calculated from reports containing forced outages excluding maintenance.

This non-standard with regard to the IEEE-762 standard summary of types and reasons was chosen because it simplifies further use in simulations.

Aggregations of the performance indicators were calculated in two different ways: capacity-weighted (w) and unweighted (u). In capacity weighting, the figures are weighted according to the rated capacity of the respective generation unit.

Since unplanned outages often only involve a reduction in output but do not require the power plant to be shut down completely, the equivalent values, which take into account both total outages and derations, are particularly useful here, in addition to the rates for total outages. For this reason all three variants have been calculated for factors and rates.

The actual measures are briefly described in **Table 5**, but a detailed overview can be found in the **supplementary materials**:

Performance measure[4]	Abbreviation	Description
Availability factor	AF	**Fraction of time without outages**
Outage hours	OH	**Cumulated hours in outages**
Outage factor	OF	**Fraction of time in outages**
Outage rate	OR	**Share of time in outage relative to operating time**
Repair rate	RR	**Average time from start to end of an outage event**
Failure rate	FR	**Event rate of an outage event**
Event counts	N	**Number of single outage events**
Mean deration	/	**Mean deration of partial outage events**
Standard deviation of derations	/	**Standard deviation of derations of partial outage events**

Table 5: Overview about calculated performance measures independent of type and reason.

Factors and rates for the specific outage types are the usual measures used to describe performance, but also for further use in simulations[25]. The repair time can be interpreted as the average duration of an event, i.e., how long maintenance takes on average in the event of an unplanned outage. The failure rate can be interpreted as the average duration between the occurrence of two outage events of a given type. Unless otherwise specified, all units describing time duration are calculated in hours. Rates and factors are relative and can be converted into percentages. Although the name suggests otherwise, the repair rate is also calculated in hours. The number of events N describes the frequency of individual events that have occurred, e.g., of type "forced". The mean value and standard deviation of a performance limitation describe the first two statistical moments of the distribution in the event of partial outages of the respective outage type and are calculated in MW. Total outages were not taken into account, as these always correspond to the nominal power and would skew the value upwards[12].

With regard to the rates for planned or unplanned outages, etc., a simplification had to be made: according to the standard, the service times of a power plant are actually required to calculate rates such as the forced outage rate (FOR). These are determined from the active times during which the power plant is active minus reserve shutdowns. Those shutdowns describe the unavailability of a power plant, e.g., due to negative electricity prices, insufficient demand, or the impossibility of operating the power plant economically and are therefore also called 'economical shutdowns'[5]. Since these reserve shutdowns are not included in the outage time series and would have to be derived with uncertainty from e.g., historical market prices, they are not taken into account. Instead, service hours were calculated as active hours minus hours with total scheduled outages from (SOH). We therefore assume that a power plant was always potentially in service when it was not scheduled to be out of service and could theoretically produce electricity.

The entire table for, e.g., 'kpis_tech_ALL.csv' contains 140 columns. An example excerpt can be seen in **Table 6**:

Column	Abbreviation	Description
Country	country	**ISO-A2 name of the country (e.g., "NO", "DE")**
ENTSO-E plant type	plant_type	**Plant type of the ENTSO-E domain (e.g., "Biomass")**
Active hours	ACTH	**Cumulated number of hours of entries for a power plant (typically a whole year)**
Service hours	SH	**Active time minus hours in total scheduled unavailability**
Observed hours	DH	**Active time minus hours in total planned unavailability**
Forced outage hours	FOH	**Cumulated number of hours during forced outages.**
Equivalent availability factor	wEAF	**Capacity-weighted fraction of time without total outages or derations**
Forced outage factor	uSOF	**Unweighted fractions of time with scheduled outages**
Unplanned deration factor	wUDF	**Capacity-weighted fractions of time in unplanned deration**
Equivalent forced outage rate	uEFOR	**Unweighted shares of time in forced outage relative to service hours**
Repair rate of forced outages	wRR_FO_h	**Capacity-weighted average time from start to end of a forced outage event (in hours)**
Failure rate of maintenance outages	uFR_MO_h	**Unweighted event rate of a maintenance outage event (in hours)**
Event counts for planned outages	N_PO	**Cumulated number of single outage events for planned outages**
Mean deration for unplanned outages	wEUD_mw_mean	**Capacity-weighted mean deration of unplanned partial outage events (in MW)**
Standard deviation of derations	uEFD_mw_std	**Unweighted standard deviation of derations of forced partial outage events (in MW)**

Table 6: Overview about calculated performance measures for the data set.

Technical Validation

Historical generation

To verify the aggregated generator outages, **Figures 4-6** exemplarily shows curves of generation and available capacity relative to installed capacity for some countries of the dataset with their thermal and/or hydraulic plant types from ENTSO-E Transparency platform and available data for generation ("AggregatedGenerationPerType_16.1.B_C_r3") as well as installed capacity ("InstalledGenerationCapacityAggregated_14.1.A_r3")[26,27]. All data were aggregated to daily averages. The information on installed capacities is published annually and is therefore sometimes very inaccurate, especially for RES, because capacity is constantly being expanded during the year. Due to the restriction that only generation units with a minimum rated power of 100 MW should be included, there should be no outage reports for individual RES plants actually. However, the graph also shows restrictions for wind offshore. If generation is greater than the specified available capacity, there may be several reasons for this. On the one

hand, the reported generation or the reported outage level could be incorrect, but on the other hand, there could also be an error in the clustering and processing of the outage reports for the specific case. The obvious caps on hard coal and natural gas in particular suggest errors in the raw data. Results for the remaining countries can be found in the **supplementary materials**.

Other publications and studies

The data set was compared with various other studies: the European Resource Adequacy Assessment (*ERAA*) 2023[28] and Ten Years Network Development Plan (*TYNDP*) 2024[29] by ENTSO-E, as well as the journal article by Gjorgiev et al[3]. A comparison with availability factors for German plant types from Bassini[4] can be found in the **supplementary materials**. Data sets are compared using absolute values of the KPIs. The *ERAA* data set was chosen because there is data modelled for the current year 2025. From the *TYNDP* study we took earliest data for 2030. With *ERAA*, we can compare forced outage rates unweighted and capacity-weighted by plant type from the ENTSO-E domain. Since the methodological reports did not provide any information on how the data was specifically determined, we compare it with FOR and UOR of our dataset, also taking forced maintenance into account. Regarding *TYNDP*, we can also compare planned outage days by converting our hourly results into days in addition to the FOR. Since it is not clear from the *TYNDP* data on planned unavailability in days why the shutdown is taking place, we also compare this with POD, SOD, and MOD. For us, MOD also includes forced outages due to maintenance, but in principle it reflects the meaning of POD in the *TYNDP* study.

For most countries and power plants, it can be seen that *ERAA* therefore assumes higher FORs (**Figure 7**). The same applies to the *TYNDP* data in **Figure 8**. One reason for the heterogeneous results of the FOR with *ERAA* as well as *TYNDP* is probably the problem described above with the derivation of service hours. Our rates tend to underestimate reality due to the simplified assumptions of economic shutdowns.

The PODs are a KPI per power plant in *TYNDP*. To make this value comparable with our aggregated annual average for all power plants of a given type, we calculate all values per plant. Since it is also unclear whether a failure day in *TYNDP* refers to a total failure or only a deration, we additionally compare the equivalent performance indicators from total failure plus derations. Because the earliest capacity expansion year from the *TYNDP* is 2030 and some power plants may have been decommissioned or newly connected to the grid by then, small differences from our historical values may slightly distort the results. It is striking that, especially for gas-fired power plants, the number of estimated days of planned outages in our data set is significantly underestimated in almost all countries compared to *TYNDP* (**Figure 9**). One reason for this could be the clustering and labelling approach (Methods – Step 3), where, in the event of a time overlap between planned and unplanned outages, the unplanned outage is selected as the label. As a result, many power plants tend to record forced outages for longer than planned outages. The differences illustrate the sometimes varying reporting practices of the TSOs with regard to the type and reason for outages: sometimes the reports with all maintenance are closer to the TYNDP value (e.g.: RO, RS or CZ) and sometimes the planned maintenance with various reasons (e.g.: FR, BE or HU).

Since the methodology used to derive forced and planned outages in the *TYNDP* 2024 does not reveal the underlying key indicators or assumptions, the expected total unavailability of a power plant within a year was also compared with the available data set. As before, all indicators are computed per single power unit of a plant type. The results can be seen in **Figure 10**.

The most comprehensive comparison can be made against the data from Gjorgiev et al., because they provide similar empirical performance measures based on the same raw data. The results for (**Figure 11**) or availability factors (**Figure 12**) are very homogeneous for most countries, with minor differences, except for individual power plant types in some countries (i.e., Biomass in Belgium or France as well as Gas in Slovakia or Bulgaria). The tendency toward higher availability in the present data set compared to the comparative data can be explained, among other things, by the calculation of a power plant's active hours, since Gjorgiev et al. counted the hours between outage reports and thus potentially did not take into account the times without outages before the first report and after the last report of their investigation horizon[3].

In **Figure 13** we also compare planned and unplanned days in outage. Since we do not know the number of power unit per plant type used by Gjorgiev et al. in their study, we sum up the values for hours in planned or forced outages (vice versa for the equivalent measures) and divide them by 24 to have an overall counter of days in planned or forced outage for a plant type. Our measures for all scheduled as well as unplanned outages or derations (including maintenance) have smaller deviations compared to just planned or forced KPIs, why we only plot them for comparison.

Usage Notes

The data can be directly used in energy system models, electricity market models, power plant dispatch models, and optimal power flow calculations, or as input in simulations to generate time series for the above-mentioned models. Average availability factors for power plant technologies or fuel types can implicitly reflect unplanned as well as maintenance outages and limit the supply side. If the attached scripts are used to determine the availability indicators of individual power plants, restrictions can also be mapped at the grid node level, allowing changes in power flow in contingency cases to be examined. The calculation of seasonal factors allows the use of availability profiles. This simplifies the modelling of hydropower technologies in particular, as the power restriction caused by outages at individual power plants is complex to implement due to cascading interdependencies of the water inflow.

Furthermore, the script 'outages_simulation.py' provides an example of code that can be used to simulate total failure events of individual power plants based on different statistical distributions and the respective failure or repair rates. In addition, partial failures can be simulated in two

variants by limiting power output. The statistical models are based on e.g. the *Antares Simulator*[30]. An example of one algorithm of the script can be found in the **supplementary materials**. Furthermore, a static or continuous multi-state Markov-Chain model using state transition probabilities could be used for the simulations[3].

Abbreviations

Example
w/u
Weighted vs. Unweighted
(E)
Equivalent
F/P/M/U/S
Abbreviation for outage type
O/D
Total outage vs. derated
Actual measure
Rate, Factor, etc.

Declarations

Data Availability

The created dataset as well as other data from different sources are publicly available. Different licences and terms of use may apply to the underlying input data from other sources than our dataset:

- Version 1.0 of the data set can be retrieved via the Zenodo repository. This link will also point to future updates: **https://doi.org/10.5281/zenodo.18998099**
- Version 0.7.1 of the open-source power plant list from *powerplantmatching* (PyPSA) can be found on Zenodo: **https://zenodo.org/records/14785651**
- Version 1.0 of the open-source power plant database from JRC can be found on Zenodo: **https://zenodo.org/record/3266807**

Code Availability

The code for processing the raw data from the ENTSO-E Transparency Platform, merging power plant databases, calculating the performance measures of the attached data set, simulating outage time series, and reproducing all figures is available as open-source code in **Zenodo**. It can be freely used and adapted, for example, to calculate additional outage KPIs or grouped aggregations based on other power plant characteristics:

- Version 1.0 of the scripts can be retrieved via the Zenodo repository. This link will also point to future updates: **https://doi.org/10.5281/zenodo.18998099**

Acknowledgements

This work was supported in part by the Helmholtz Association under the project "**Helmholtz platform for the design of robust energy systems and raw material supply"** (RESUR) [grant number:**37.12.02**] as well as the **Helmholtz Program Energy System Design** (ESD) [grant number:**37.12.03**].

Author contributions

E.J.: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **A.A.:** Conceptualization, Writing – review & editing, Supervision.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. VGBE-TW-103V. *Availability of Power Plants 2014 – 2023*. (vgbe energy service GmbH, Essen, 2024).
2. ENTSO-E Transparency Platform. UnavailabilityOfGenerationUnits_15.1.A_B. *Transparency Platform HelpDesk* https://transparencyplatform.zendesk.com/hc/en-us/articles/40477405498257-UnavailabilityOfGenerationUnits-15-1-A-B (2025).
3. Gjorgiev, B., Stankovski, A., Wengler, J., Sencan, S. & Sansavini, G. Availability of the European power system assets: What we learn from data? *Reliab. Eng. Syst. Saf.* **258**, (2025).
4. Bassini, C. F. Seasonal and weather influences in the outages of German thermal generators. *ACM SIGEnergy Energy Inform. Rev.* **5**, (2025).
5. IEEE SA. *762-2023 - IEEE Standard Definitions for Use in Reporting Electric Generating Unit Reliability, Availability, and Productivity*. (IEEE, New York, NY, 2023).
6. N-SIDE. *Study on the Outage Parameters of Generation Units and DC Links*. (2022).
7. Pecinovsky, J. & Boerman, F. entsoe-py. (2025).
8. ENTSO-E Transparency Platform. Planned Unavailability & Changes in Actual Availability of Generation & Production Units [15.1.A] & [15.1.B] & [15.1.C] & [15.1.D]. *Transparency Platform HelpDesk* https://transparencyplatform.zendesk.com/hc/en-us/articles/16652173943828-Planned-Unavailability-Changes-in-Actual-Availability-of-Generation-Production-Units-15-1-A-15-1-B-15-1-C-15-1-D (2025).
9. ENTSO-E Transparency Platform. New Transparency Platform Website Go Live. *News* https://transparency.entsoe.eu/news (2025).
10. ENTSO-E Transparency Platform. Unavailability of Production and Generation Units [15.1.A&B&C&D]. *Transparency Platform HelpDesk* https://transparencyplatform.zendesk.com/hc/en-us/articles/32492060810513-UnavailabilityOfProductionAndGenerationUnits-15-1-A-B-C-D-r3 (2025).
11. NERC. *Generating Availability Data System - Data Reporting Instructions*. (2025).
12. North American Electric Reliability Corporation (NERC). Derate Event Reporting - Data Reporting Instructions - Section III. (2025).
13. North American Electric Reliability Corporation (NERC). Outage Event Reporting - Data Reporting Instructions - Section III. (2025).
14. Schmitz, R., Frischmuth, F., Braun, M. & Härtel, P. Coping with Risk Factors in Energy System Transformations - Climate Change Impacts on Nuclear Power Plant Availability in Europe. in *2024 20th International Conference on the European Energy Market (EEM)* (IEEE, Istanbul, Turkiye, 2024). doi:10.1109/EEM60825.2024.10608936.
15. Finck, J. R. Techno-economic assessment of market coupling regimes in future electricity systems. (Karlsruher Institut für Technologie, Karlsruhe, Deutschland, 2024).
16. Gils, H. C., Bothor, S., Genoese, M. & Cao, K.-K. Future security of power supply in Germany - The role of stochastic power plant outages and intermittent generation. *Int. J. Energy Res.* **42**, (2018).
17. Deakin, M., Greenwood, D., Brayshaw, D. J. & Bloomfield, H. Comparing Generator Unavailability Models with Empirical Distributions from Open Energy Datasets. in *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (IEEE, Manchester, United Kingdom, 2022). doi:10.1109/PMAPS53380.2022.9810629.
18. Kanellopoulos, K., De Felice, M., Hidalgo Gonzalez, I. & Bocin, A. JRC Open Power Plants Database (JRC-PPDB-OPEN). Zenodo https://doi.org/10.5281/ZENODO.3266807 (2019).
19. Gotzens, F., Heinrichs, H., Hörsch, J. & Hofmann, F. Performing energy modelling exercises in a transparent way - The issue of data quality in power plant databases. Zenodo https://doi.org/10.5281/ZENODO.14785651 (2025).
20. Sergio, A. & Colelli, F. P. Weather-induced power plant outages: Empirical evidence from hydro and thermal generators in Europe. *Energy Econ.* **148**, 108549 (2025).
21. Murphy, S., Sowell, F. & Apt, J. A time-dependent model of generator failures and recoveries captures correlated events and quantifies temperature dependence. *Appl. Energy* **253**, (2019).
22. Ersayin, E. & Ozgener, L. Performance analysis of combined cycle power plants: A case study. *Renew. Sustain. Energy Rev.* **43**, 832–842 (2015).
23. Guénand, Y. *et al.* Climate change impact on nuclear power outages - Part I: A methodology to estimate hydro-thermic environmental constraints on power generation. *Energy* **307**, (2024).
24. Collet, L. *et al.* Future nuclear power outages in a changing climate - A case study on two contrasted French power plants. *Energy* **320**, (2025).
25. Curley, G. M. Reliability Analysis of Power Plant Unit Outage Problems. (2013).
26. ENTSO-E Transparency Platform. Actual Generation per Production Type [16.1.B&C]. *Transparency Platform HelpDesk* https://transparencyplatform.zendesk.com/hc/en-us/articles/16648290299284-Actual-Generation-per-Production-Type-16-1-B-C (2025).
27. ENTSO-E Transparency Platform. Installed Generation Capacity Aggregated [14.1.A]. *Transparency Platform HelpDesk* https://transparencyplatform.zendesk.com/hc/en-us/articles/16648300912916-Installed-Generation-Capacity-Aggregated-14-1-A#:~:text=A%5D,-2%20months%20ago&text=The%20sum%20of%20installed%20Net,generation%20capacity%2C%20per%20production%20type. (2025).
28. ENTSO-E. *European Resource Adequacy Assessment 2023 Edition - Annex 1: Input Data and Assumptions*. https://www.entsoe.eu/eraa/2023/report/ERAA_2023_Annex_1_Assumptions.pdf (2024).
29. ENTSO-E. *TYNDP 2024 Scenarios Methodology Report - Final Version*. https://2024.entsos-tyndp-scenarios.eu (2025).

30. Antares-Simulator Team. Antares-Simulator. *Antares Simulator Documentation* https://antares-simulator.readthedocs.io/en/latest/user-guide/ts-generator/05-algorithm/ (2025).

Footnotes

1. As it turned out that there were still duplicates of the identification codes, the line with the lowest available capacity was selected to cover the worst-case scenario.
2. The difference of 2,536 entries is due to the fact that additional countries were filtered and the start and end dates of the report must be between January 2015 and October 2025.
3. Northern Ireland is included in Great Britain. Iceland, Malta, Ukraine, the Azores, the Canary Islands, the Balearic Islands, and other overseas territories of France or Great Britain were not included.
4. The term “outage” in this list includes not only total outages, but also performance limitations and equivalents consisting of total and partial outages.

Figures

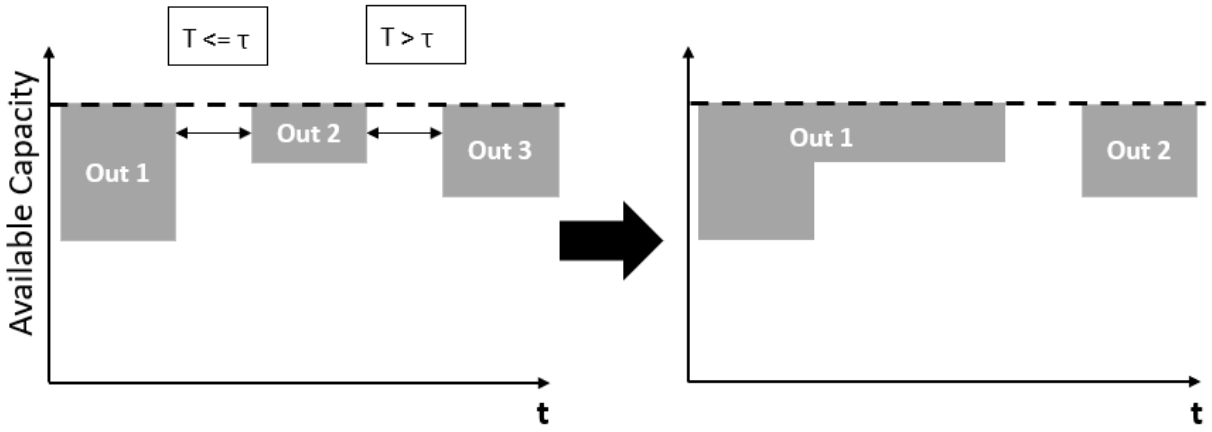


Figure 1

Method to bridge gaps between outage reports using a defined maximum temporal difference.

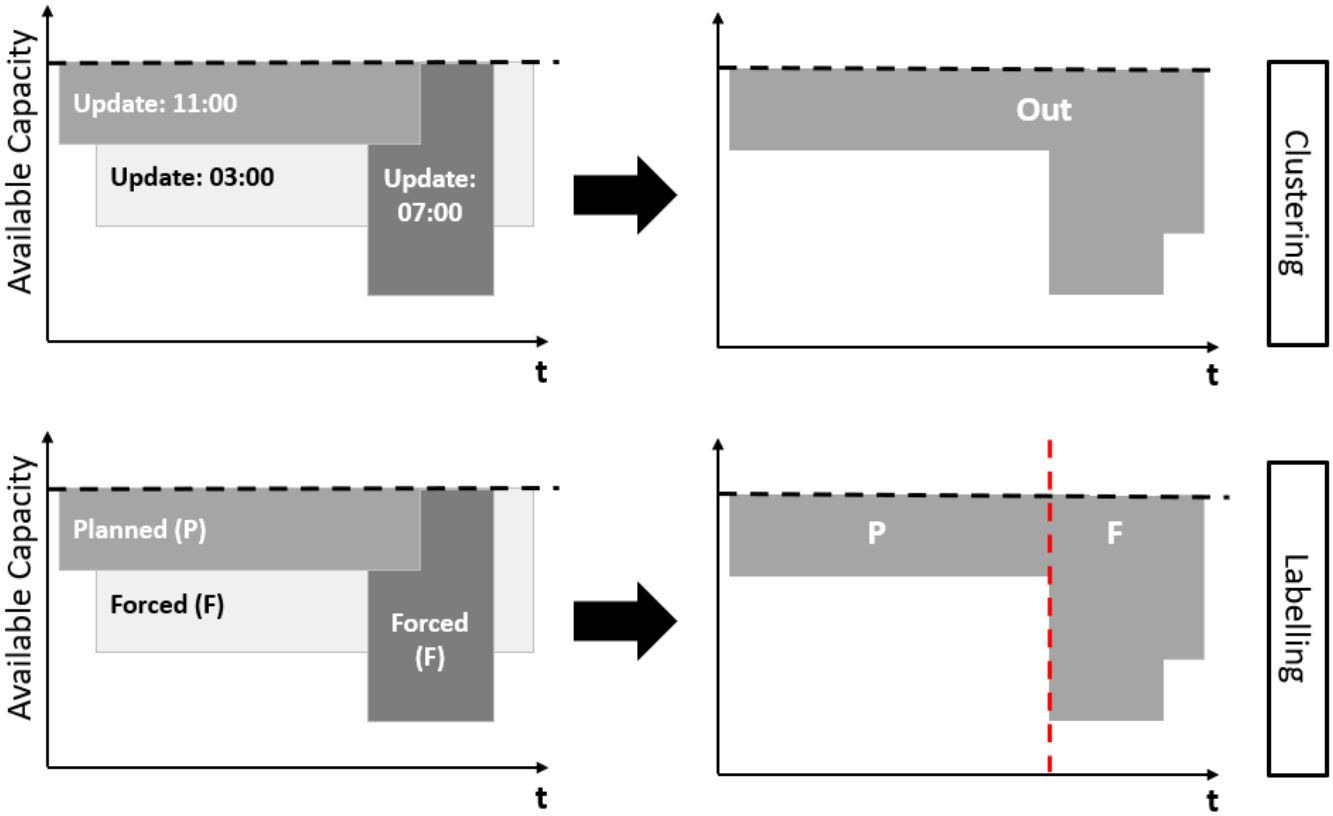


Figure 2

Clustering and labelling of multiple different overlapping outage reports.

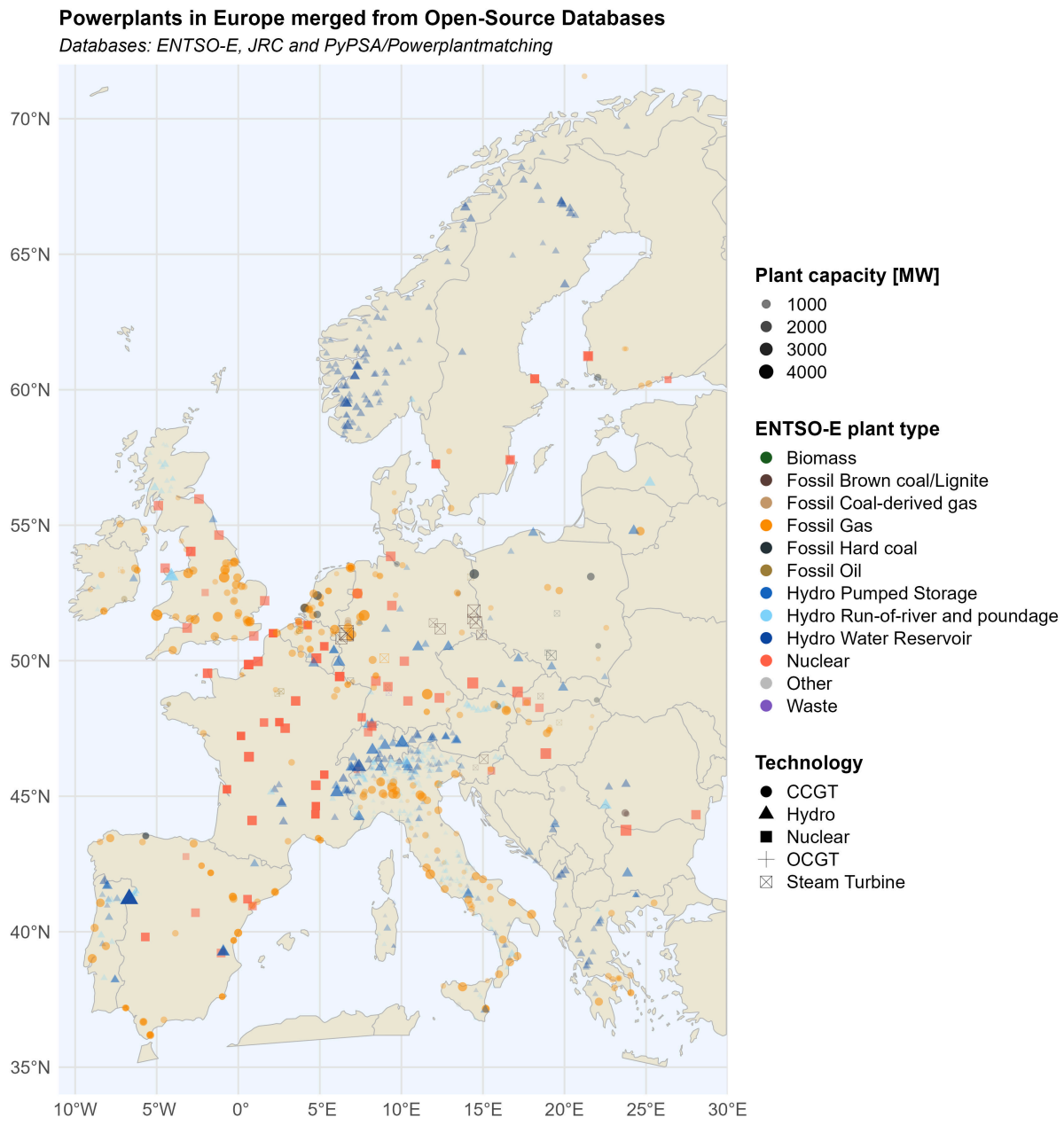


Figure 3

Map of the power plant list used for the statistical analysis merged by databases from ENTSO-E, JRC and PyPSA.

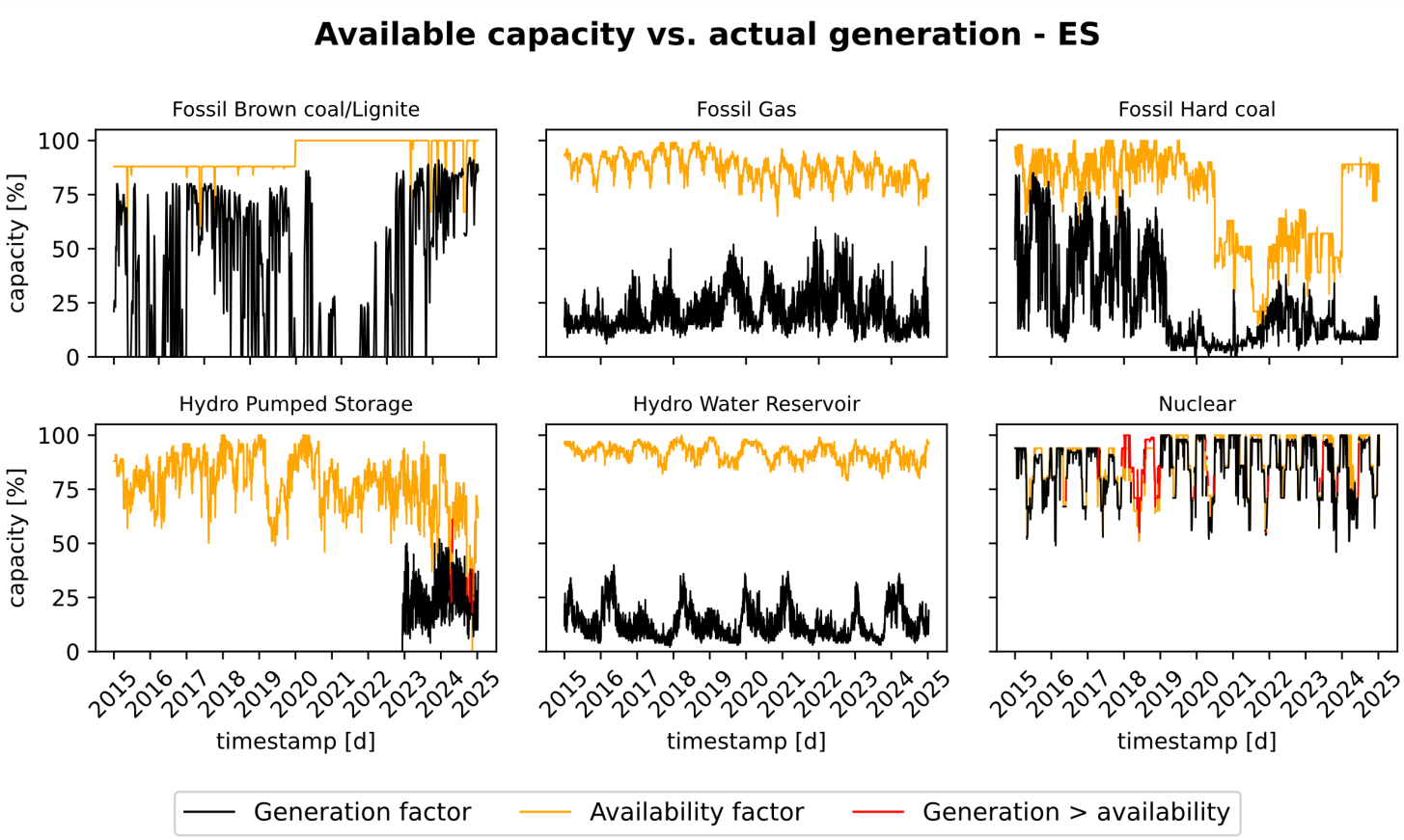


Figure 4

Comparison of aggregated available capacity vs daily generation relative to installed capacity for Spain.

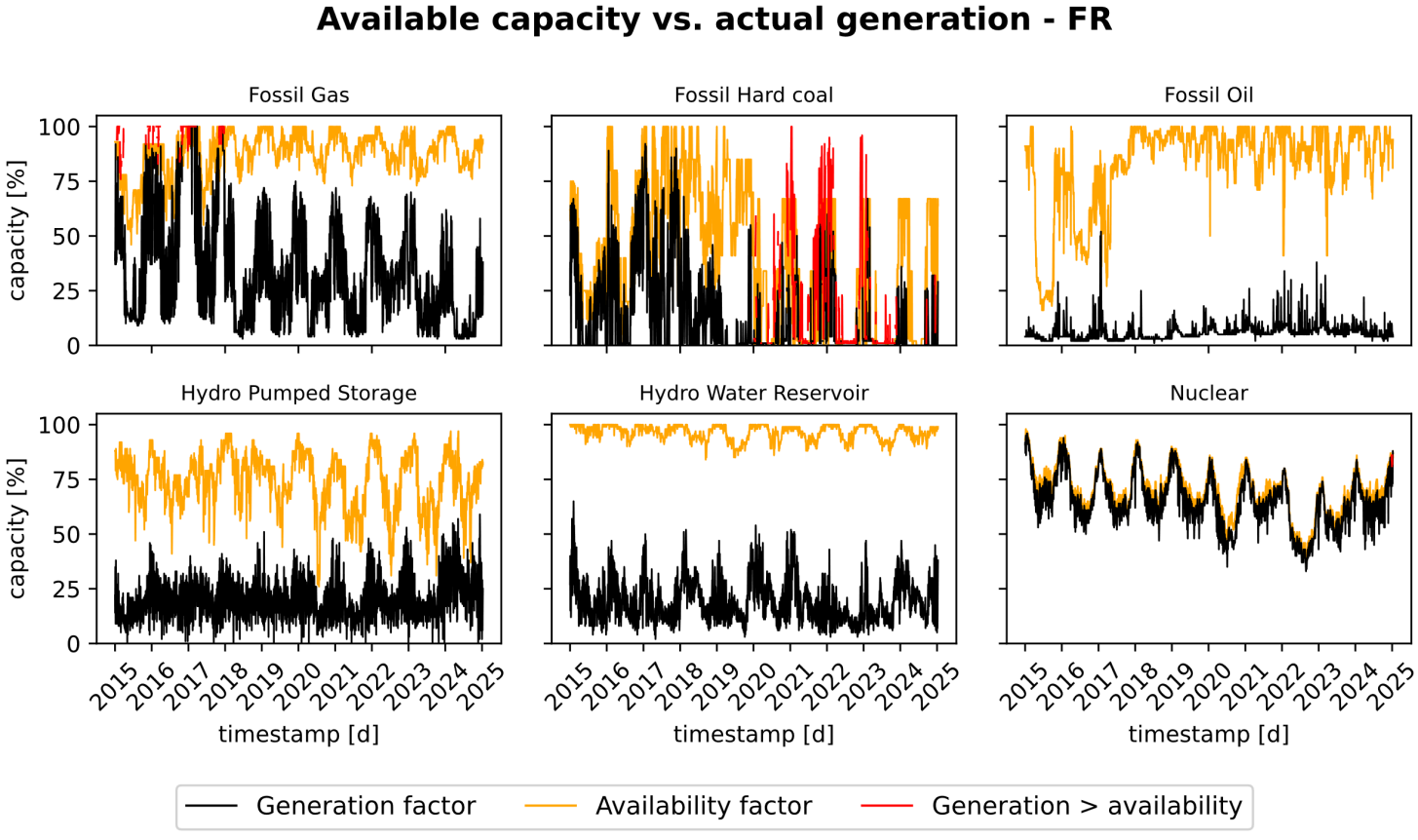


Figure 5

Comparison of aggregated available capacity vs daily generation relative to installed capacity for France.

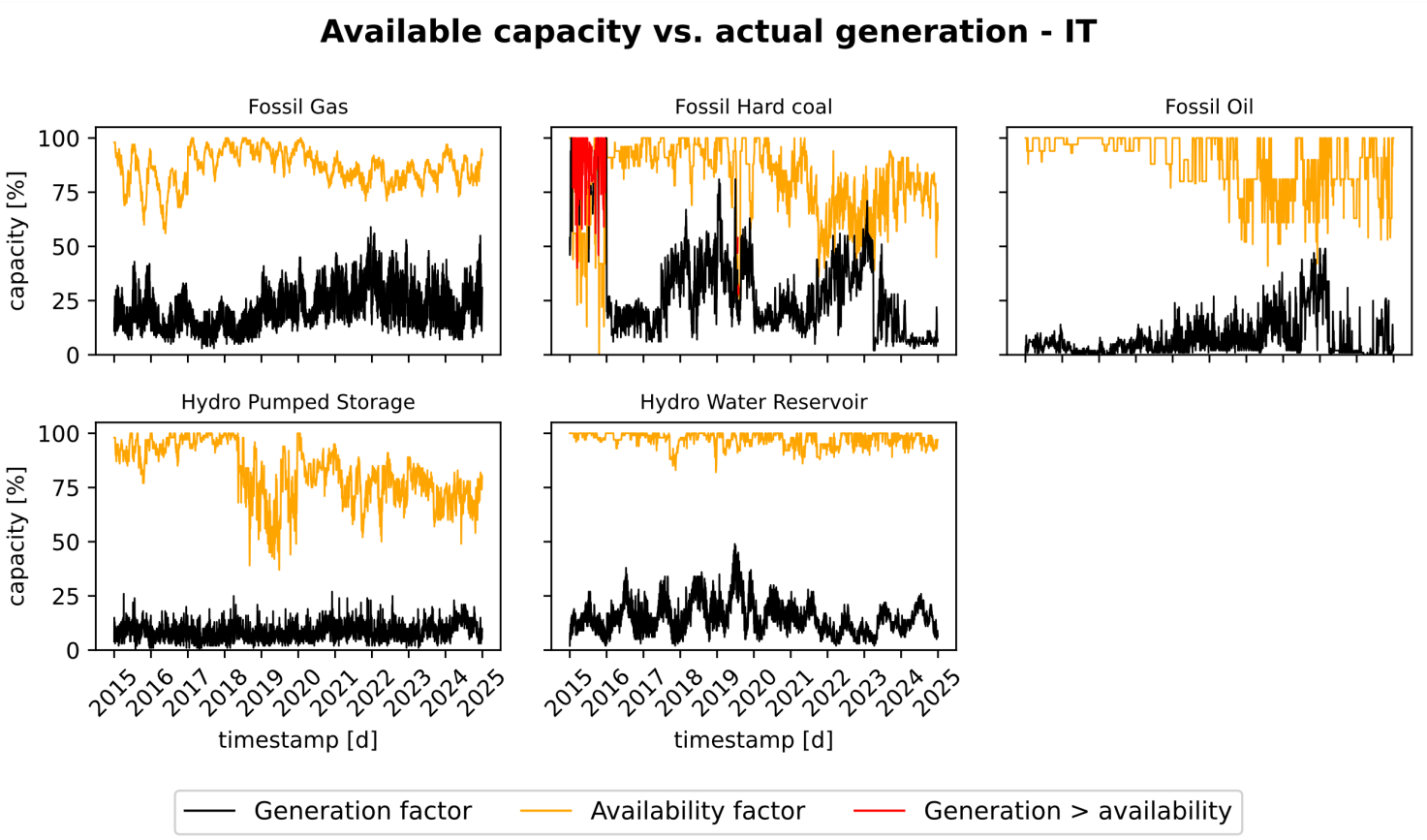


Figure 6

Comparison of aggregated available capacity vs daily generation relative to installed capacity for Italy.

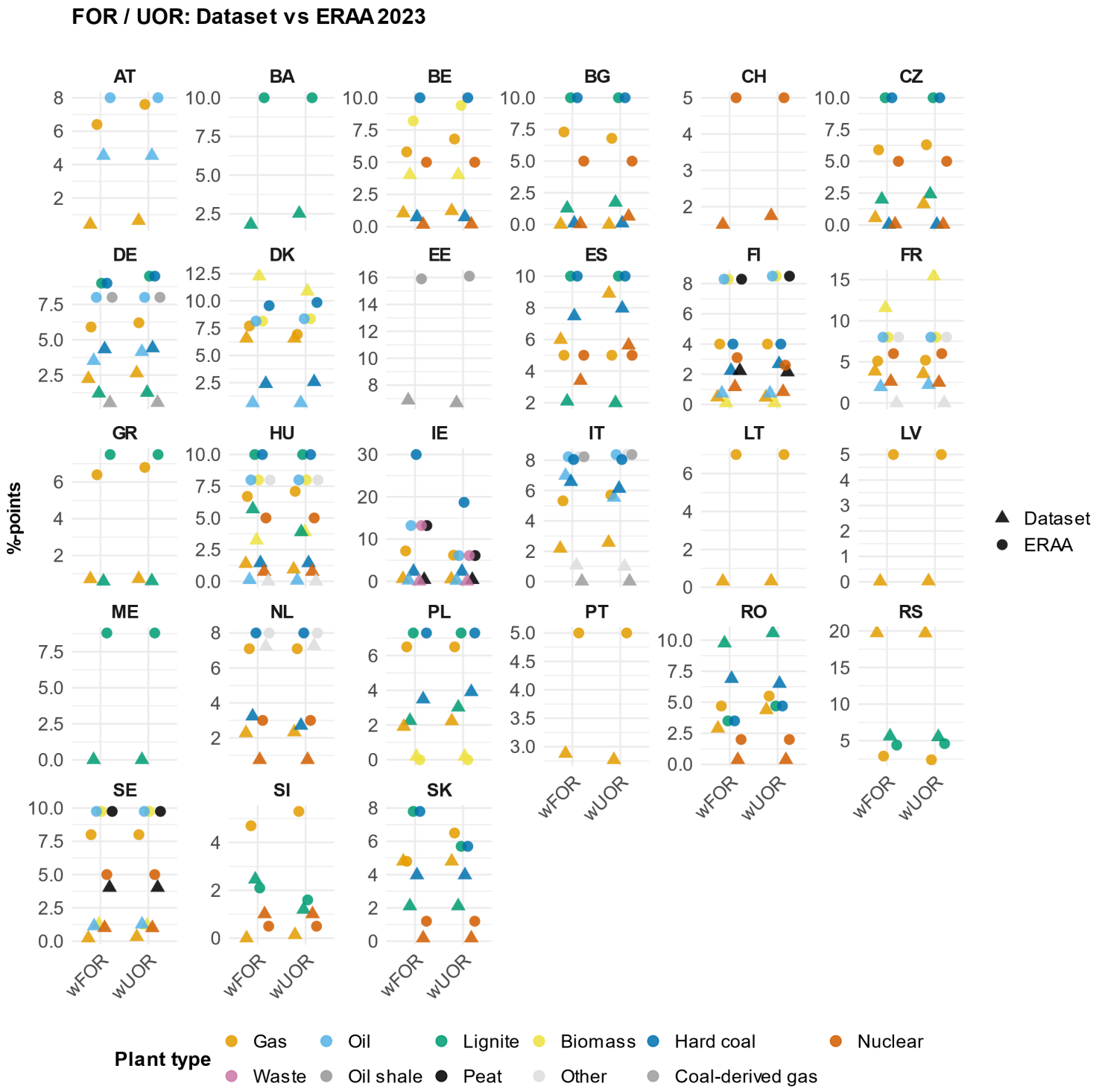


Figure 7

Comparison of FOR from ERAA 2023 against FOR and UOR from our dataset.

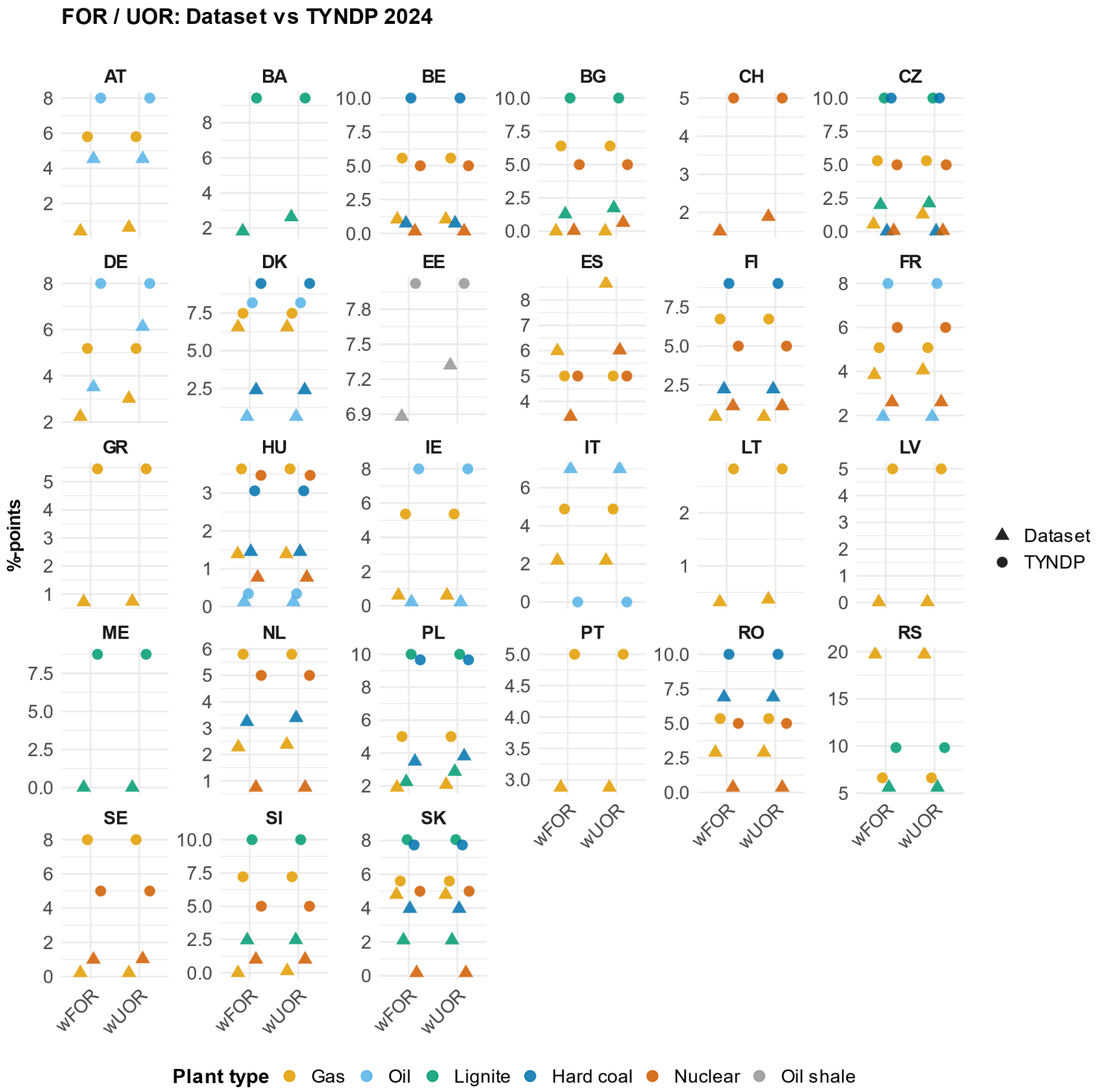


Figure 8

Comparison of FOR from TYNDP 2024 against FOR and UOR from our dataset.

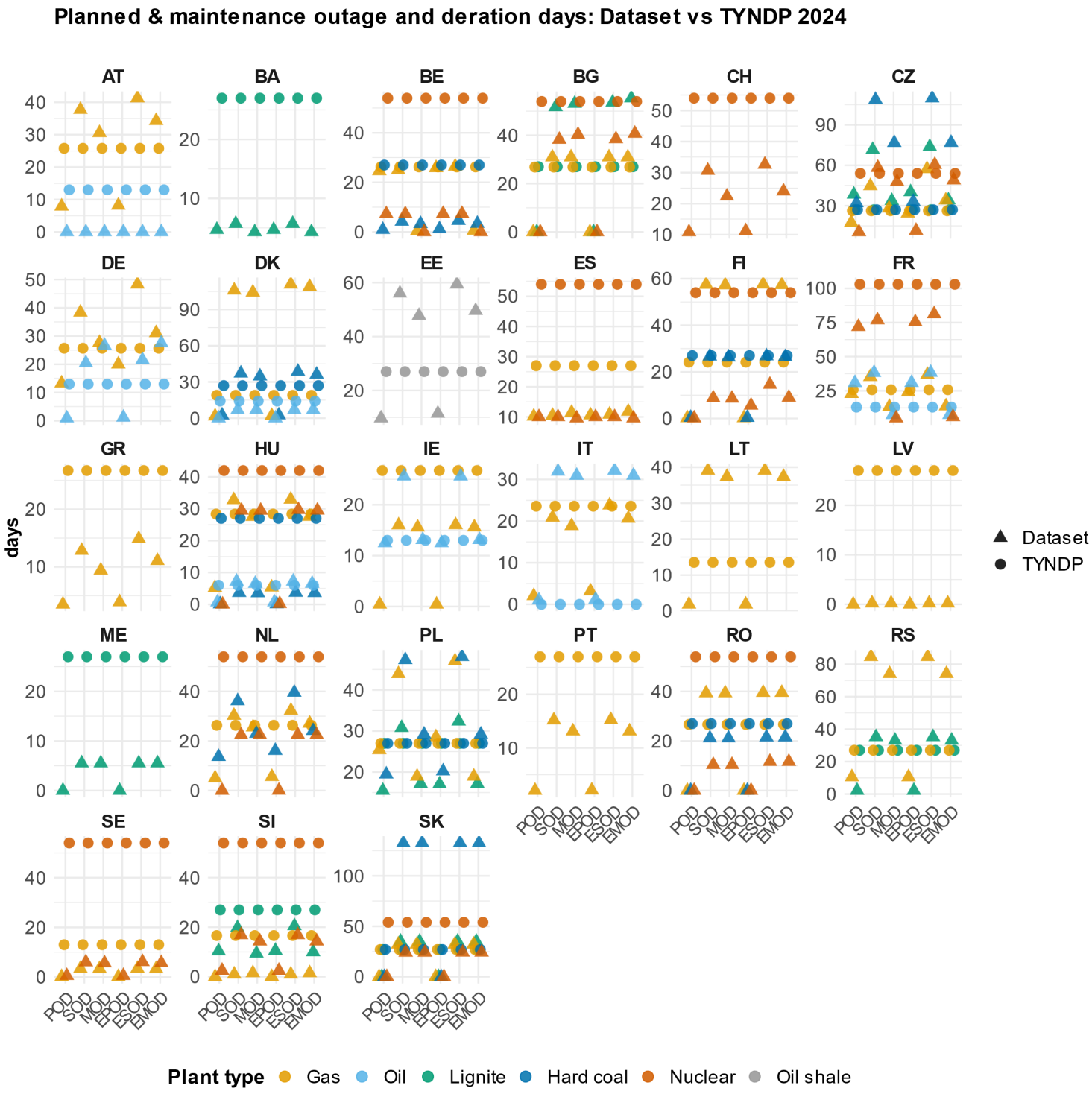


Figure 9

Comparison of POD from TYNDP 2024 against planned and maintenance outages as well as derations from our dataset.

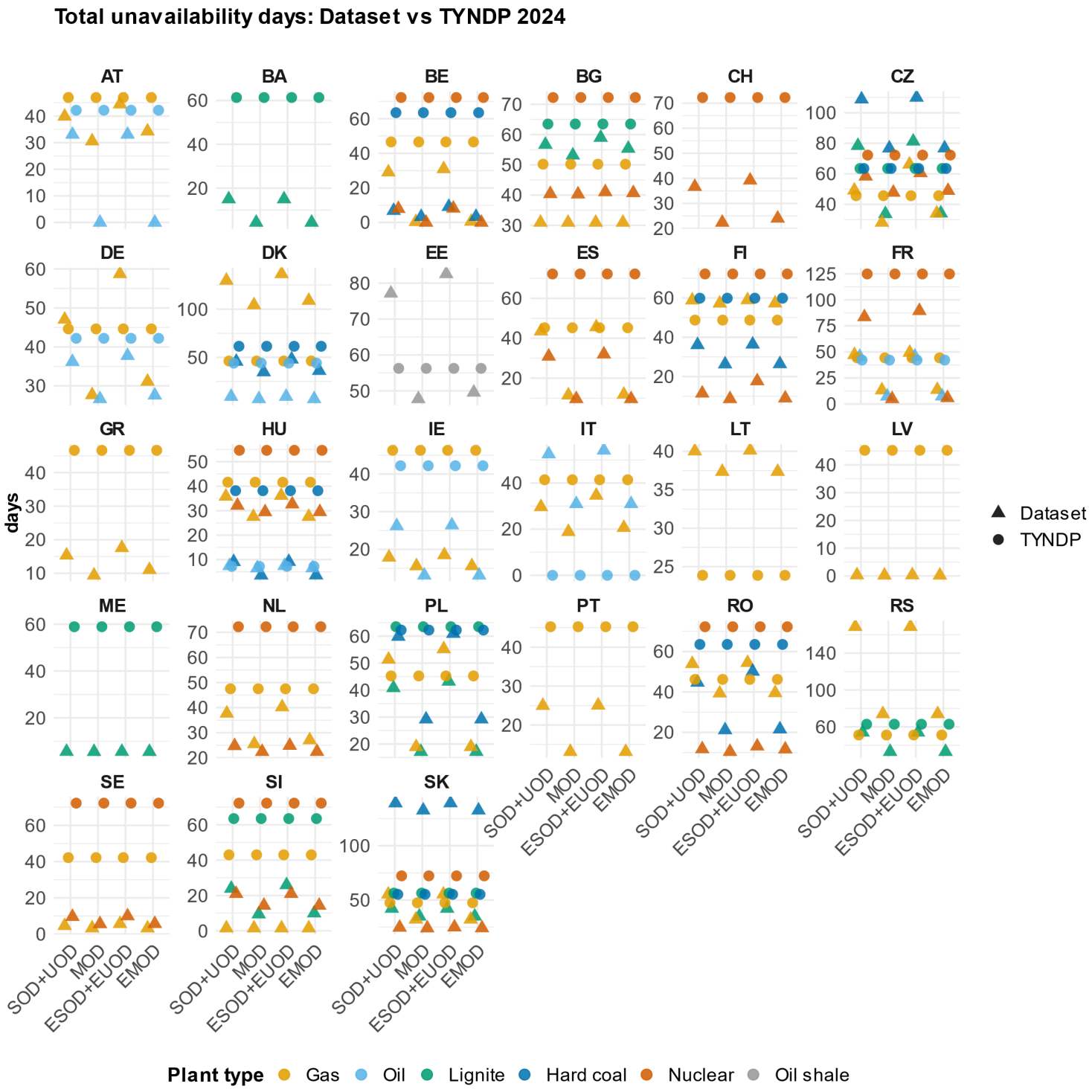


Figure 10

Comparison of expected total unavailability during a year per power unit of a given type from TYNDP 2024 against our dataset.

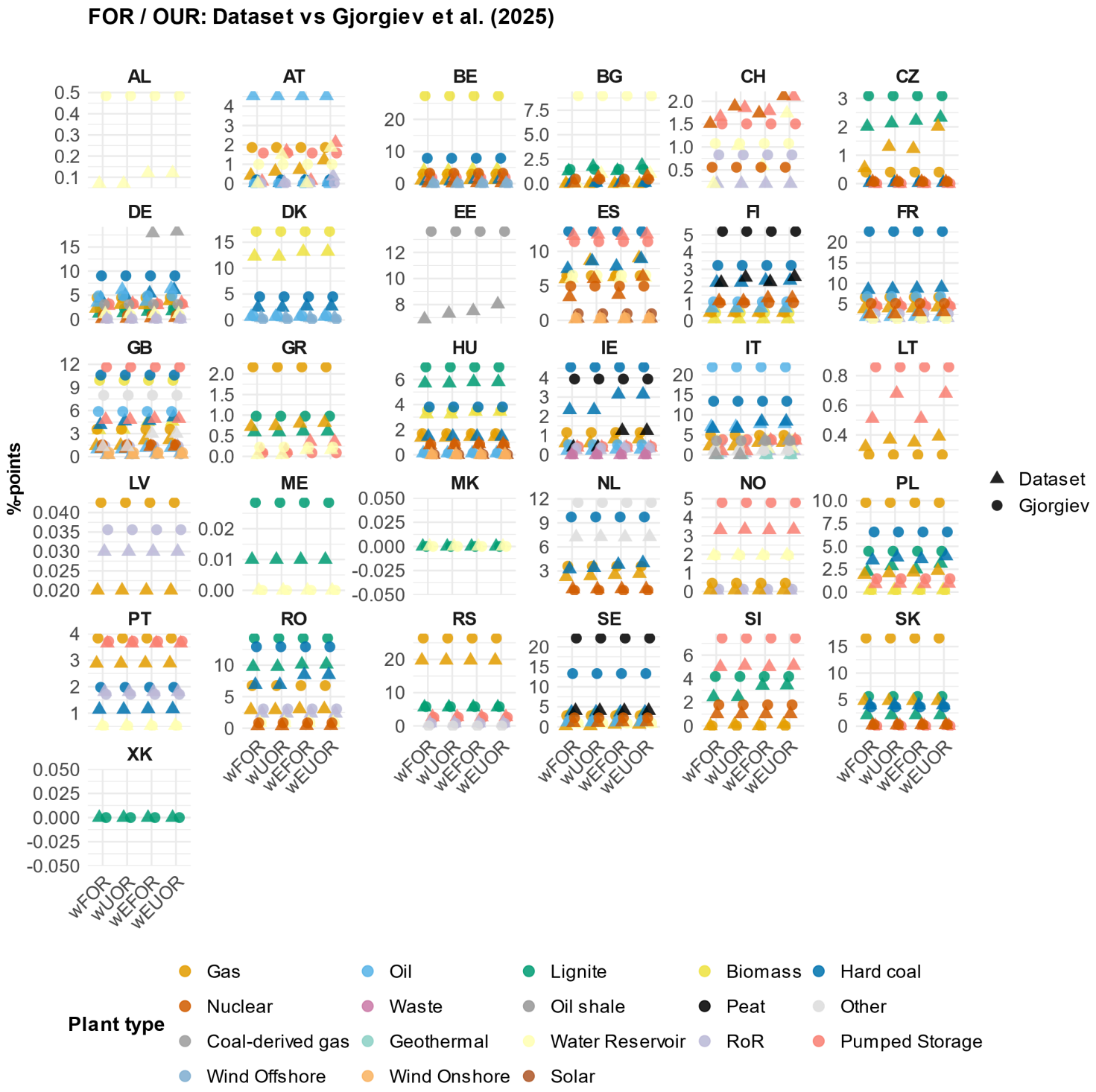


Figure 11

Comparison of FOR and UOR with Gjorgiev at al. 2025

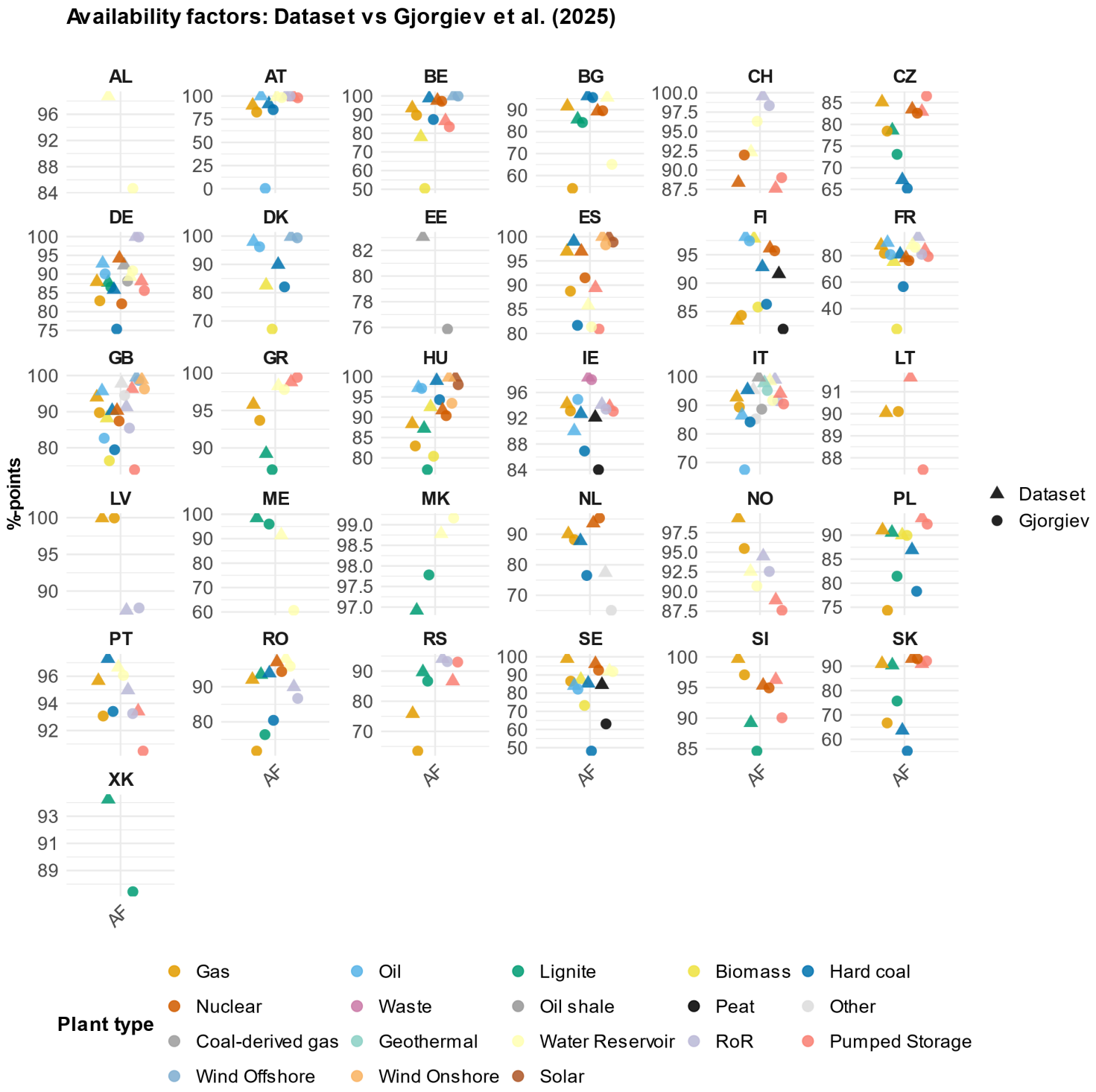


Figure 12

Comparison of availability factors with Gjorgiev et al. 2025

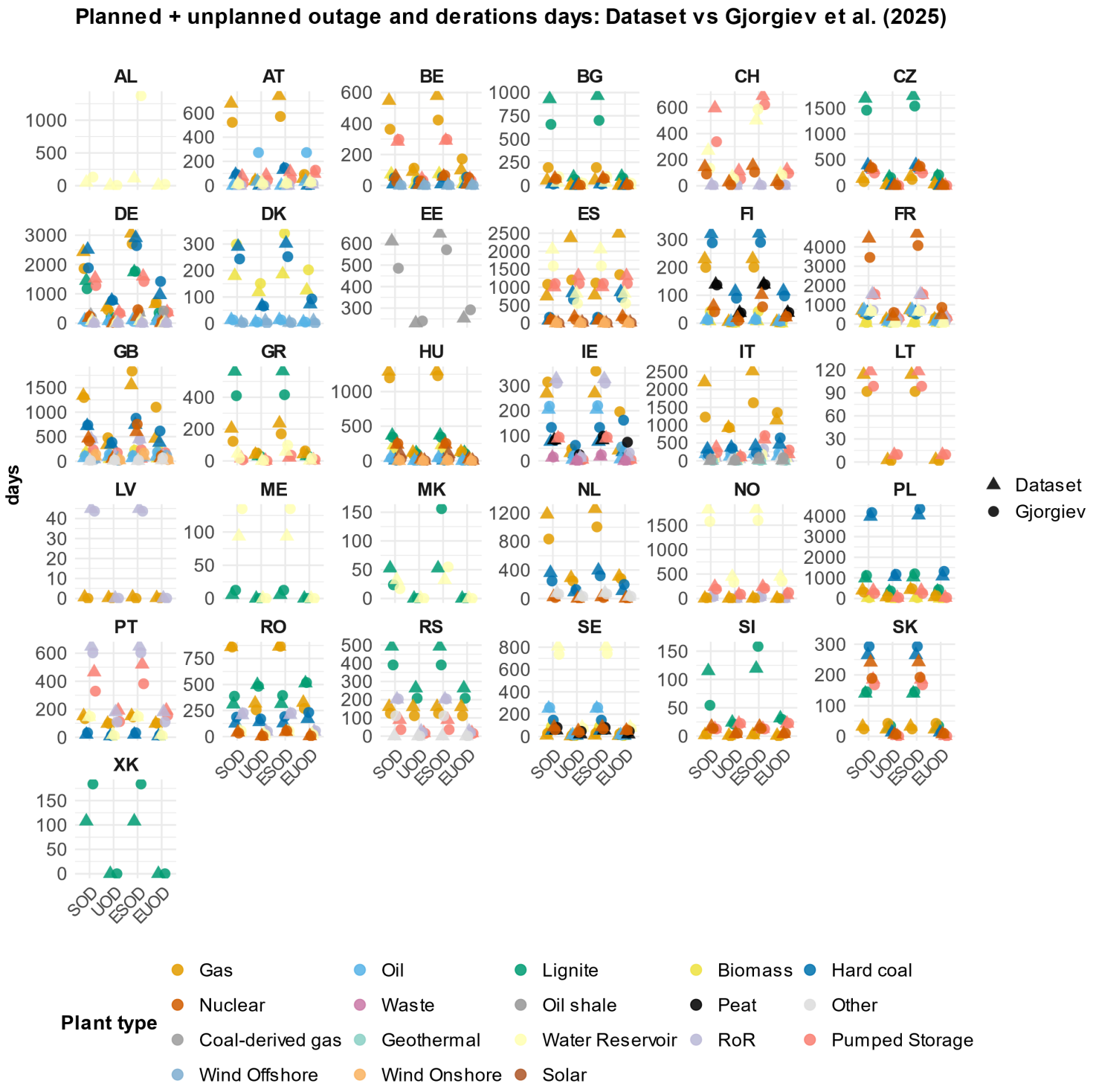


Figure 13

Comparison of planned as well as unplanned outage and derations days with Gjorgiev et al. 2025

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx