Power System Modelling in the Baltic Countries: Data Accessibility and Consistency Aspects

Jana Teremranova, Zane Broka, Kārlis Baltputnis
Institute of Power Engineering
Riga Technical University
Riga, Latvia
jana.teremranova@rtu.lv, zane.broka@rtu.lv, karlis.baltputnis@rtu.lv

Abstract—**In this paper we present a review of the data necessary for Baltic power system and electricity market modelling, analyse the most important and useful data sources, their accessibility, quality and utility. We provide observations regarding the studied data sources as well as highlight and address some of the most pressing issues that arise in this domain, including data consistency between various sources. Based on a data comparison example concerning power generation timeseries in Latvia, we show how researchers need to be cautious when selecting the databases and information sources to be used in their modelling efforts. In the end, lessons learned are summarized and recommendations made to both data providers and users.**

Keywords—electricity market; open science; open data; power generation; power system modelling

I. Introduction

As European policymakers, power system operators and other energy industry stakeholders are developing the regulatory framework and increasing the technical readiness to be able to adapt to an increasingly growing role of intermittent renewable energy sources, there are also discussions on the need and possible actions towards a European electricity market reform [1]. While the calls for a reform are a more recent trend driven by the electricity price shock experienced in 2021 and 2022, the path to increasingly more renewable energy in the power system has a longer history [2]. Nevertheless, the ambition in that regard is also growing with calls for a fully renewable power system more and more often taking the central stage [2].

Power system and electricity market modelling is an instrumental tool towards designing and validating mechanisms for achieving both the above-mentioned goals. However, for impactful modelling, input data quality is of paramount importance. Moreover, to allow for the reuse of the modelling tools and reproduction of the results, it is important to ensure that data is accessible and properly licensed. Unfortunately, in practice, licensing issues are often neglected by power system modellers [3]. However, the awareness of the importance of open data as a crucial part of the open science process has been increasing.

While developing an open source electricity market research model for the Baltic states [4], which includes indicators regarding Latvia (LV), Lithuania (LT), Estonia (EE), and for some parameters also the neighbouring countries like Finland (FI), Sweden (Southernmost bidding area, SE4) and Poland (PL), we have also noted that special attention needs to be paid to data issues. To this end, we have undertaken a detailed review of the data necessary for the Baltic power system and electricity market modelling, the most important data sources, their quality and utility. In the current paper, we provide observations regarding data sources useful specifically for modelling the Baltic states as well as highlight and address some of the most important issues that arise in this regard.

The remainder of the paper is structured as follows. Section II summarizes the data needed for power system and electricity market modelling and identifies the most prospective data sources, providing also assessment of accessibility issues. Section III addresses data consistency considerations whereby, based on a selected example, it is shown how data values of the same indicator can differ across sources. Finally, the paper is concluded with lessons learned and recommendations addressed to both data users and providers.

II. Data Availability

The main indicators of interest and the spatial and temporal requirements placed on them for use in model development and validation are depicted in Table I.

TABLE I. Initial Spatial and Temporal Requirements for Baltic Electricity Market Modelling Data

Indicator	Spatial scope	Time resolution
Generation capacity	LT, LV, EE	yearly
Generation by source type		hourly
Generation by large units		
Power plant unavailability		
Electricity consumption		
Interconnection capacity and unavailability	FI-EE, EE-LV, LV-LT, LT-SE4, LT-PL	
Energy stored in hydro reservoirs	LV, LT	weekly
Water inflow to hydro reservoirs	LV	hourly / as available
Price and cost data:		
Wholesale electricity market	LT, LV, EE, FI, SE4, PL	hourly
CO_2 emission allowances	EU	daily / monthly
Natural gas	LT, LV, EE (or EU)	weekly / monthly
Biomass and biogas		
Shale oil		
Heavy fuel oil		

This research is funded by the Latvian Council of Science, project "Multi-functional modelling tool for the significantly altering future electricity markets and their development (SignAture)", project No. lzp-2021/1-0227.

As mentioned in [4], for effective use and sharing of an open-source model, it is very important to have access to data that is transparent, openly available and regularly updated. At least hourly resolution is required wherever possible. Our interest is generally in a five-year horizon of historical data to monitor the dynamics of indicators. However, the timespan can be increased if necessary for other models and purposes.

A. *ENTSO-E Transparency Platform*

Electricity generation is one of the main parameters required for power system modelling (i.e., for calibrating and validating models, and for preparing non-dispatchable generation timeseries), and it is available from various sources. One of the largest and most significant is the data repository of the European Network of Transmission System Operators for Electricity, ENTSO-E Transparency Platform [5] (hereinafter – Platform), which offers access to a very thorough and continuously updated dataset of all the European power systems. Therein, 39 Transmission System Operators (TSOs) from 35 countries gather and upload data of electricity load, generation, transmission, balancing, outages, etc., providing both recent and historical data with (mostly) hourly resolution. The aggregated generation per production type (16.1.B&C) timeseries for Lithuania is available from 5 January 2015; for Latvia since 24 December 2014; and for Estonia since 17 December 2014. While it has some access limitations when viewing the platform via the web interface: only data for one day is available for viewing at once, it is however possible to download the hourly data for a full year in *xml*, *csv* and *xlsx* format from the visual interface as well. Moreover, for automatic data extraction, it is possible to use the application programming interface (API). This is true for all the data available on the Platform.

Generation data per production type can also be obtained from other sources, such as Eurostat [6] and national statistics portals ([7] for LT, [8] for LV and [9] for EE). However, hourly timeseries are not available there, instead offering the aggregate monthly and annual data.

Information on energy stored in water reservoirs and pumped storage plants is also available on the Platform, albeit with weekly resolution (16.1.D). There are the aggregated stored energy values for Latvia (which has a cascade of three large hydropower plants (HPP) on the Daugava River) and Lithuania (which has the Kruonis pumped storage plant). There is no notable water reservoir capacity in Estonia. Regrettably, the Platform does not have data regarding water inflow to the Daugava River, which is an important parameter for more accurate modelling of the operation of the largest HPPs in Latvia.

There are also data entries concerning the installed capacities (14.1.B), generation timeseries (16.1.A) as well as outages and unavailability of individual production and generation units (15.1.A-D). However, this data is only available for large units, i.e., those above 100 MW of installed capacity, which is the threshold set by the respective regulation [10]. Unfortunately, the interpretation of what constitutes unit capacity (i.e., if it should be considered at individual generator or plant level) differs [11]. E.g., for Latvia there are only five generation units with individual timeseries data on the Platform. Namely, the Platform contains timeseries of Riga CHP-2 (combined heat and power plant) gas and steam turbine generation in both units as well as a single unit (generator No. 4) of Plavinas HPP, which is part of the Daugava HPP cascade consisting of 23 hydropower units in total.

In terms of the licensing of the data published on the Platform, ENTSO-E provides a list of freely reusable data [12] which is licensed under Creative Commons Attribution 4.0 International License (CC-BY 4.0) and hence does not require the "need to seek for the prior agreement of the respective Primary Owner of Data.". However, the list does not include all the indicators necessary and useful for power system modelling efforts. For the remaining items, licensing issues are unclear and are subject to due diligence of the data user requiring time-consuming investigation to clarify the terms of licensing for each specific dataset and data provider.

B. *Open Power System Data*

Another source of electricity data timeseries available for use and download is the Open Power System Data (OPSD) platform [13]. This platform collects, aggregates and publishes data from various open-source repositories, allowing to obtain in an easy-to-handle form well-structured data necessary for researchers involved in modelling energy systems. *xlsx*, *csv* and *sql* data formats are available for downloading in both manual and machine-readable ways. Information is aggregated on the OPSD platform both from international (e.g., ENTSO-E, Eurostat) and national (e.g., national statistics, TSOs, DSOs) sources.

However, as concerns the Baltic power system data, the OPSD only has timeseries entries regarding wind generation, solar generation and actual and forecasted load for Lithuania, Latvia and Estonia as well as annual installed capacities in each country. The platform also has temperature and radiation data, which, while not highlighted in Table I, is nevertheless a potentially useful asset, especially if the modelling work is focused more towards studying intricacies related to PV generation development.

C. *Nord Pool*

Traditionally, a major source of electricity generation, consumption and price data is the electricity market operators. Nord Pool is of primary interest for the Baltic states as it operates the day-ahead and intraday markets there. However, since the beginning of 2022, historical data is no longer freely accessible on the site [14] and thus it is less attractive for open-source modelling following the FAIR (findable, accessible, interoperable, reusable) principles.

D. *Open-Source Project Databases*

Finally, a wholly different approach to data curation for modelling is re-using databases aggregated in other projects and modelling efforts. In the Baltic states context, the most notable of databases with free access stem from the BENTE (*Baltic Energy Technology Scenarios*) [15] and FASTEN (*Fast, flexible, and secure decarbonization of the Baltic states – possible progress in the next Ten years*) projects [16].

E. Data Source Summary

An overview of the main parameters for modelling of the Baltic power system and electricity market and their prospective sources are summarized in Table II.

TABLE II. PROSPECTIVE DATA SOURCES FOR BALTIC ELECTRICITY MARKET MODELLING

Indicator	Source
Generation capacity	[5], [7]–[9], [13], [15], [17]
Generation by source type	[5]–[9], [13], [14], [18]–[20]
Generation by large units	[5], [15], [17]–[20]
Power plant unavailability	[5]
Electricity consumption	[5]–[9], [14], [18]–[20]
Interconnection capacity and unavailability	[5], [14], [19]
Energy stored in hydro reservoirs	[5]
Price and cost data:	
Wholesale electricity market	[5], [14], [19]–[21]
CO_2 emission allowances	[22]
Natural gas	[19], [20], [23]
Biomass and biogas	[15], [17], [24]
Shale oil	[15], [17]
Heavy fuel oil	[15], [17]

As can be discerned from the table, the primary source for price and cost statistics is generally the market operator of the respective exchange (for electricity, gas, biomass etc.). However, electricity market price can be generally accessed from TSO-related sources, including the ENTSO-E Transparency Platform.

On the other hand, there are several cost indicators where the identification and use of a primary source is less clear (e.g., shale oil, heavy fuel oil, biogas). For these indicators, it is rational to base assumptions on prior open-source projects.

III. DATA COMPARISON CASE STUDY AND RESULTS

The use of data for power system modelling implies trust in the data source. In this section, based on selected examples, we explore the consistency and comparability of data obtained from different sources. For the sake of brevity, these examples are limited in geographical scope to Latvia and only concern the electricity generation domain.

First, we compare generation data from different sources. We can notice significant differences between the hourly generation by type timeseries when comparing the data obtainable from the Latvian TSO, Augstsprieguma tikls AS (AST), website [18] and the corresponding data published on the ENTSO-E Transparency Platform [5]

Fig. 1 illustrates the hourly difference between these timeseries in 2022, calculated by summing up the generation of all types of sources in each hour and comparing them by relating the difference between the ENTSO-E and AST data to the ENTSO-E value per hour. Evidently, there are many hours with very high relative differences in values ranging from +82.05% to –388.89%. The largest differences occur mostly in hours when electricity production in absolute numbers is rather low in Latvia, thereby the relative variation of values between the two sources can be more pronounced.

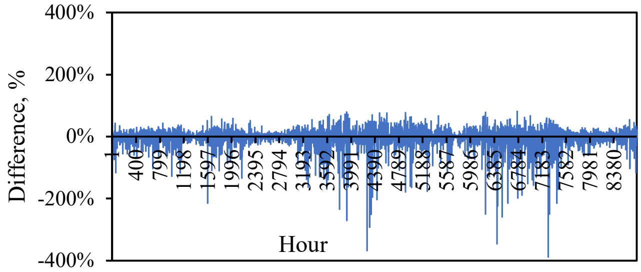

Fig. 1. Electricity generation hourly timeseries percentage difference between ENTSO-E [5] and national TSO [18] data for the year 2022

Interestingly, however, the annual average relative difference of the hourly values is only –1.35% (or –1.63 MWh/h in absolute terms). In 50.49% of hours, the generation values are higher in ENTSO-E data; in 44.36%, there is larger generation in TSO data; in 5.15% of hours, the numbers are the same.

In the relevant section on the AST website, there is a disclaimer that these are untested and unverified data, but it is unclear if this disclaimer is also applicable to the historical datasets (available from 2013) or only to the current-day view.

By randomly examining the numerical values in more detail, we found that the supposed historical hourly generation values from the respective TSO data table (*State of Latvian power system / Scheduled and actual consumption, net exchange and production*) actually coincide with the instantaneous value on the 1st minute of each hour available in the more detailed table (*State of Baltic-Nordic Power system / Production types, consumption and net exchange*). It follows that the values in question published by the TSO indeed are not actually the hourly energy production (MWh/h) but rather the instantaneous power (MW) at the top of the hour.

Unfortunately, the tables with minute resolution have less utility in modelling since it is possible to download data for only one day in one pass and there is no API access on the AST website, which thus potentially mandates the use of an automated data scraping script to obtain the necessary data for much longer periods.

Thereby, after comparing the two specific sources for hourly generation timeseries, it is evident that ENTSO-E data prevails as potentially more trustworthy as we can assume these data have been validated and corrected after the initial publication as necessary. The next question to then explore is the quality of the timeseries uploaded to ENTSO-E. For this, we compare the generation timeseries on the Platform to those published by the market operator Nord Pool.

Fig. 2 reflects the difference between ENTSO-E and Nord Pool electricity generation data for Latvia in 2020. The year 2020 has been selected because it was the last year fully available before Nord Pool closed the open access to its data. Here the situation looks much better than in the previous comparison. Except for two outlier peaks (+6.9% and –4.1%) around the mid-year most likely caused by some technical issues, the spread of hourly differences is small, and therefore both sources of information can be considered fairly equal.

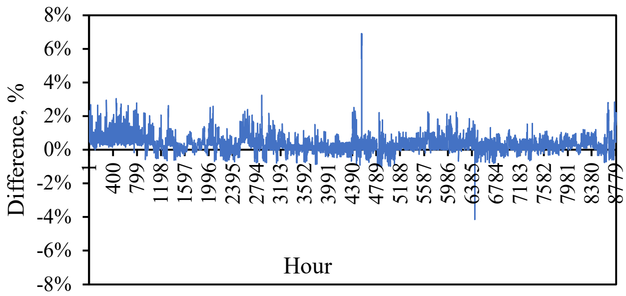

Fig. 2. Electricity generation hourly timeseries percentage difference between ENTSO-E [5] and Nord Pool [14] data for the year 2020

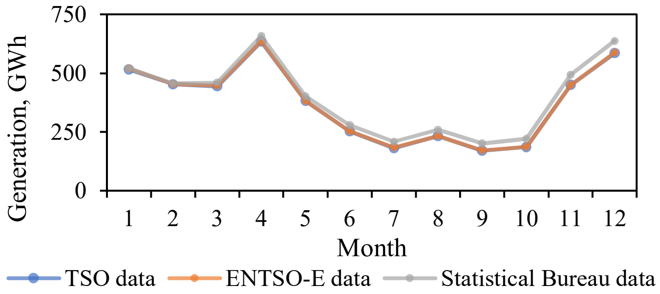

Fig. 3. Monthly electricity generation in Latvia in 2022 according to different data sources

There is an apparent trend for generation values in the ENTSO-E database to be larger compared to Nord Pool, which is the case in 67% of hours in 2020. However, in 24% of hours there is no difference between both datasets. Hence, it is hard to explain the variance with systemic issues (e.g., one of the datasets including additional considerations like self-consumption), so it could rather be a case of one of the datasets being validated/corrected more recently than the other one. Nevertheless, the absolute difference between these datasets is not significant, on average constituting +1.83 MWh/h in 2020.

Finally, we conclude the data comparison by analysing the difference of monthly generation in 2022 (Table III). We use the hourly data from the Latvian TSO (AST) and ENTSO-E, aggregated per month, and compare it with the monthly generation published on the official Latvian statistics portal [8].

TABLE III. MONTHLY ELECTRICITY GENERATION IN LATVIA IN 2022 FROM DIFFERENT DATA PLATFORMS

Month	TSO [18], GWh	ENTSO-E [5], GWh	St. Bureau [8], GWh	Difference [18] & [5]	Difference [18] & [8]
Jan	515	519	520	–0.75 %	–0.91 %
Feb	453	453	456	–0.15 %	–0.75 %
Mar	444	447	460	–0.78 %	–3.60 %
Apr	635	637	658	–0.30 %	–3.66 %
May	382	382	402	0.04 %	–5.12 %
Jun	252	253	279	–0.07 %	–10.52 %
Jul	180	184	208	–1.83 %	–15.31 %
Aug	233	233	259	–0.28 %	–11.38 %
Sep	170	171	201	–0.81 %	–18.19 %
Oct	185	186	220	–0.36 %	–18.64 %
Nov	449	450	493	–0.02 %	–9.68 %
Dec	585	588	637	–0.39 %	–8.84 %

Notably, the difference between the TSO and ENTSO-E data does not exceed 2%. The TSO reports slightly lesser production in all the months except for May. On the other hand, the differences with the data from the Central Statistical Bureau of Latvia are much more significant reaching up to 18.64%, with the production values always being lower in the TSO database.

These values are also summarized in Fig. 3. Note that the line representing TSO data is not visible in the chart as it is barely different from the ENTSO-E line. Quite evidently, the differences are the smallest towards the beginning of the year. This might imply that the datasets are being corrected with a delay of more than a year as the timeseries concern 2022, whereas the data for comparison was retrieved in mid-2023.

All things considered, different approaches to collecting the data on electricity generation in Latvia and the subsequent data publishing can be found, namely:

- The hourly aggregated generation on the ENTSO-E Transparency Platform is the average of all available instantaneous net generation values during the hour (or respective market time unit). If net generation is partially or fully unavailable, it is estimated. Generation of small power plants is estimated if real-time measurement data are not available [5].
- Nord Pool indicates that for Latvia the published data are net generation values provided by the TSO, 94% of which are measured and 6% are estimated data (this concerns mainly small generation units as the information about their electricity production is obtained by the TSO at a later stage but seemingly is not updated on Nord Pool website) [25].
- The Latvian TSO (AST) asserts that the hourly data published on its webpage dashboard is only illustrative, unverified and not guaranteed to be accurate [18]. Verified monthly data is assured to be available in the TSO's monthly electricity market reviews and Nord Pool's website. However, the TSO's market reviews do not provide granular hourly data, nor is the generation timeseries freely available from Nord Pool anymore, as indicated previously.
- The monthly generation data by source published by the Statistical Bureau of Latvia concerns gross electricity production in CHPs, HPPs, wind and solar power plants and thus includes self-consumption and losses, i.e., the monthly data per generation type are available only as gross generation. For net production, only the monthly aggregate value is provided without distinguishing generation type [8].

IV. CONCLUSIONS AND RECOMMENDATIONS

Overall, there is a large number of sources where the data necessary for electricity market model calibration and validation can be found. However, they can differ greatly in terms of data granularity, quality, accessibility, consistency, documentation, and licensing conditions among other aspects. To ensure that sufficiently reliable data is used for modelling, it is recommended to cross-check the datasets between different databases whenever possible. One such exercise is shown in this paper as an example.

Considering the issues highlighted above and still existing at the time of preparing this paper, for hourly generation data we recommend relying on the ENTSO-E Transparency Platform. As concerns power generation in Latvia, we cannot recommend energy modellers to use the historical generation timeseries available on the online dashboard of the Latvian TSO (AST) for modelling at hourly granularity as significant differences could be found when comparing them with the Transparency Platform. However, it can still serve as an illustrative tool for observing the dynamics of the electricity market and the Latvian power system. Moreover, the dashboard data could be especially useful when instantaneous generation values with sub-hourly granularity are of interest for more detailed modelling of power system dynamics.

Even though the Latvian TSO does warn about the non-validated nature of the dashboard archive data, one could recommend the TSO to continuously update the historical dataset with validated timeseries as they become available. There should also be a clear distinction between the unverified operational data and validated historical datasets. Moreover, the ease-of-use when downloading the data should be improved and access via API enabled. When it comes to monthly data, however, the data provided by the Latvian TSO can be treated as more reliable. Nevertheless, as shown in the study, there are significant differences between some of the sources, which, however, decrease the further in the past we look. This shows that the monthly data is being verified gradually, thus the differences between the data sources become increasingly insignificant over time.

All things considered, ENTSO-E could be viewed as the superior historical data source because of the vast number of indicators available in hourly resolution and the varied access options. However, there are also some drawbacks over which we recommend the ENTSO-E to carry out further work to improve the quality, ease-of-use and shareability of data. First, there are occasional issues with inexplicably missing data points in the ENTSO-E Transparency Platform, which sometimes even differ depending on if the data is accessed through the web interface or via API.

Second, apart from the timeseries explicitly listed by ENTSO-E as freely useable under the CC-BY 4.0 license, the data redistribution options are subject to agreements with primary data owners, which often enough are not clearly identifiable.

In its turn, the OPSD platform aims to reduce the inconveniences arising from data quality and licensing issues by processing missing values and aggregating freely available data useful for modelling efforts in one place. Unfortunately, as of now there is very limited data there regarding the Baltic power system. To this end, it is recommended for energy modellers as well as for primary data owners to more actively engage in open data initiatives.

In continuation of this work, further studies and data analysis can be carried out on the consistency of electricity generation data per type, generation capacities, electricity consumption etc. depending on the specific modelling needs by comparing the data provided by different sources, potentially also expanding the geographical scope.

REFERENCES

[1] Leonardo Meeus and Sofia Nicolai, "A summary of the proposal for a reform of the EU electricity market," *Florence School of Regulation*, Mar. 17, 2023. https://fsr.eui.eu/a-summary-of-the-proposal-for-a-reform-of-the-eu-electricity-market/ (accessed Aug. 02, 2023).

[2] K. Hansen, C. Breyer, and H. Lund, "Status and perspectives on 100% renewable energy systems," *Energy*, vol. 175, pp. 471–480, May 2019, doi: 10.1016/j.energy.2019.03.092.

[3] L. Hirth, "Open data for electricity modeling: Legal aspects," *Energy Strategy Rev.*, vol. 27, p. 100433, Jan. 2020, doi: 10.1016/j.esr.2019.100433.

[4] Z. Broka and K. Baltputnis, "Open-source electricity market modelling for the Baltic states: review and requirements," in *2023 19th International Conference on the European Energy Market (EEM)*, Jun. 2023, pp. 1–6. doi: 10.1109/EEM58374.2023.10161953.

[5] ENTSO-E, "ENTSO-E Transparency Platform." https://transparency.entsoe.eu/ (accessed Mar. 14, 2023).

[6] Eurostat, "Energy Database." https://ec.europa.eu/eurostat/web/energy/database

[7] Statistics Lithuania. State Data Agency, "Official Statistics Portal." https://osp.stat.gov.lt/

[8] Central Statistical Bureau of Latvia, "National statistical System of Latvia." https://data.stat.gov.lv/pxweb/en/

[9] Statistics Estonia, "Statistical Database." https://andmed.stat.ee/en/stat

[10] *Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009 of the European Parliament and of the Council Text with EEA relevance*. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2013/543/oj

[11] "A review of the ENTSO-E Transparency Platform," VVA, Copenhagen Economics, Neon, Deloitte, 2017. [Online]. Available: https://energy.ec.europa.eu/system/files/2018-05/review_of_the_entso_e_plattform_0.pdf

[12] ENTSO-E, "List od Data Available for Free Re-use." Feb. 18, 2022. [Online]. Available: https://transparency.entsoe.eu/content/static_content/download?path=/Static%20content/terms%20and%20conditions/220218_List_of_Data_available_for_reuse.pdf

[13] "Open Power System Data." https://open-power-system-data.org/

[14] Nord Pool AS, "New Data Portal for 2022." https://www.nordpoolgroup.com/en/trading/Operational-Message-List/2022/01/new-data-portal-for-2022-20220114124600/

[15] T. J. Lindroos *et al.*, *Baltic Energy Technology Scenarios 2018*, 2018:515. in TemaNord. Copenhagen: Nordic Council of Ministers, 2018. doi: 10.6027/TN2018-515.

[16] N. Putkonen *et al.*, "Modeling the Baltic countries' Green Transition and Desynchronization from the Russian Electricity Grid," *Int. J. Sustain. Energy Plan. Manag.*, vol. 34, pp. 45–62, May 2022, doi: 10.54337/ijsepm.7059.

[17] "Backbone / projects / FasTen-model · GitLab," *GitLab*, Oct. 27, 2021. https://gitlab.vtt.fi/backbone/projects/fasten-model (accessed Aug. 01, 2023).

[18] Augstsprieguma tīkls AS, "Power System state." https://ast.lv/en/content/power-system-state (accessed Mar. 14, 2023).

[19] Litgrid, "Dashboard." https://www.litgrid.eu/index.php/dashboard/630 (accessed Aug. 01, 2023).

[20] Elering, "Elering LIVE." https://dashboard.elering.ee/en

[21] Ember, "Data." https://ember-climate.org/data/

[22] European Energy Exchange AG, "EU ETS Auctions." https://www.eex.com/en/market-data/environmentals/eu-ets-auctions

[23] UAB GET Baltic, "Trading Data." https://www.getbaltic.com/en/market-data/trading-data/

[24] UAB Baltpool, "Biomass exchange." https://www.baltpool.eu/en/biomass-exchange/

[25] Nord Pool, "Production." https://www.nordpoolgroup.com/en/Market-data1/Power-system-data/Production1

This is a post-print of a paper published in Proceedings of the 2023 IEEE 64th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (**RTUCON 2023**) and is subject to IEEE copyright.
https://doi.org/10.1109/RTUCON60080.2023.10412855
Electronic ISBN: 979-8-3503-1773-2.
USB ISBN: 979-8-3503-1772-5.
Print on Demand (PoD) ISBN: 979-8-3503-1774-9.