

# Simulation of day-ahead electricity market prices using a statistically calibrated structural model

Valentin Mahler, Robin Girard, Sébastien Billeau, Georges Kariniotakis

► **To cite this version:**

Valentin Mahler, Robin Girard, Sébastien Billeau, Georges Kariniotakis. Simulation of day-ahead electricity market prices using a statistically calibrated structural model. 16th European Energy Market Conference (EEM 2019), IEEE PES, Sep 2019, Ljubljana, Slovenia. hal-02351932

**HAL Id: hal-02351932**

**<https://hal.archives-ouvertes.fr/hal-02351932>**

Submitted on 6 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Simulation of day-ahead electricity market prices using a statistically calibrated structural model

Valentin Mahler<sup>\*,†</sup>, Robin Girard<sup>\*</sup>, Sébastien Billeau<sup>\*</sup>, Georges Kariniotakis<sup>\*</sup>

<sup>\*</sup>MINES ParisTech, PSL University, PERSEE – Centre for Processes, Renewable Energies and Energy Systems, CS 10207 rue Claude Daunesse 06904 Sophia Antipolis Cedex, France

<sup>†</sup>ADEME – Agence de l'Environnement et de la Maîtrise de l'Energie, 20 avenue du Grésillé 49004 BP 90406 Angers Cedex 01, France

**Abstract** — Anticipating electricity prices on the day-ahead market has become a key issue for both risk assessment and revenue optimization. In this paper, we propose to generate time series of prices with an hourly resolution using a structural model that simulates a simplified market clearing process. The aggregated supply curves in this model are composed of orders based on the available capacity of generation units. The ask prices are parametrized, and the calibration is performed by applying statistical learning to historical market and power system data. To reflect the strategic behavior of market participants, these prices depend on the scarcity of power at the national level. The model's performance is assessed based on the case of France with a one-year horizon and data from 2013-2015. This approach illustrates how open data on the electric power system enable links to be drawn between technical constraints and price formation.

**Index Terms** — day-ahead markets, electricity prices, statistical learning, structural model

## NOMENCLATURE

The key notations used in this paper are introduced below.

### Indices and sets

$i$	Index of generation unit in the set $I$
$t$	Index for time step in the set $T$
$\mathfrak{t}$	Index for production type in the set $\mathfrak{T}$
$X^*$	The training period
$\tilde{X}$	A simulated price

### Variables

$\mathcal{L}_t^{net}$	Total power consumption in the bidding area covered by its dispatchable generation units at time $t$
$\mathcal{D}_t$	Demand on the day-ahead market at time $t$
$\mathcal{M}_t$	Reserve margin at time $t$
$\bar{P}_{i,t}$	Offered power supply for unit $i$ at time $t$
$P_{i,t}$	Accepted power supply for unit $i$ at time $t$
$\tilde{\lambda}_{i,t}$	Price per unit of energy for unit $i$ at time $t$
$\tilde{\Lambda}_{i,t}$	Market clearing price at time $t$
$\mathcal{L}_t^{all}$	Total electrical load in the bidding area at time $t$

$\mathcal{P}_t^{ico}$	Total net energy balance of the international interconnections in the bidding area at time $t$
$\mathcal{P}_t^{stor}$	Total net storage of the international interconnections in the bidding area at time $t$
$\mathcal{P}_t^{VRE}$	Total net energy balance of power sources considered non-dispatchable in the bidding area at time $t$
$\bar{C}_{i,t}$	Generation capacity of unit $i$ at time $t$

## I. INTRODUCTION

In the European Union and other deregulated electricity markets, wholesale financial transactions occur between large energy producers and consumers. In complement to bilateral agreements, power exchanges provide popular alternatives to trade energy, such as on day-ahead markets. These markets are typically structured so that orders to sell or buy electricity for a specific hour of the following day are transmitted by market participants to a market operator who determines a single market clearing price and the accepted orders. By trading products with a high time resolution and close to delivery, it is possible to adjust the orders based on the latest available information, such as forecasts for demand and for the production of intermittent renewable energy sources. Moreover, a liquid day-ahead market features a large pool of market participants, thus increasing the likelihood of finding a counterpart for transactions. Notably due to these advantages, power exchanges play a central role in electricity trading in Europe, and the number of reported transactions continues to grow [1, p. 2].

Understanding price formation mechanisms is crucial for companies that want to anticipate their operating costs or revenues, as well as for policy makers and energy regulators that devise and enforce appropriate regulations. Highlighting the link between the technical constraints of the electric power system and prices helps to define effective long-term investment strategies and to plan the operation of assets in the short to mid runs. Compared to other commodities, electricity is subject to specific constraints that directly impact its price dynamics. These influencing constraints include limited grid-scale energy storability, congestion of transmission lines, low elasticity of demand, and a strict demand-supply equilibrium.

The authors wish to thank the French Environment and Energy Management Agency (ADEME), the *Association pour la Recherche et le Développement des Méthodes et Processus Industriels* (ARMINES) and Corusant SA for financially supporting this research.

Mathematical models are developed to assess the quantitative impact of various factors on prices. Such models can be used within their validity domain to generate price forecasts. The simulated values ultimately provide a support for decision processes.

## II. OBJECTIVES

This necessity to anticipate price dynamics led us to develop a model designed to generate time series with an hourly price resolution on a day-ahead market. Using a statistical learning approach made possible by open data, our aim is to reveal links between fundamental data and the observed day-ahead prices. Since we simulate a market clearing price based on parametrized supply curves, the proposed model is classified as a structural model according to the definition provided in [2].

Our model is one of the few to have been developed with the goal of generating long-term forecasts [3]. The approach of basing producers' offer prices on market clearing prices has been explored in [4]. In addition, we use a scarcity factor similar to [5] in order to account for strategic bidding. Complementary approaches to infer aggregated supply curves [6] and electricity suppliers' cost functions [7] have also been proposed recently.

The proposed modeling method and a case study are presented hereafter.

## III. METHOD

### A. Overview of the method

For each time step (i.e. hour), we calculate a residual demand ( $\mathcal{D}_t$ ) and simulate offers for the power supply consisting in pairs of energy quantities ( $\bar{P}_{i,t}$ ) and price per unit of energy ( $\tilde{\lambda}_{i,t}$ ). The operations leading to these values are described in the subsections "Demand simulation" and "Supply simulation".

Using the residual demands and offers for power supply, the following optimization problem is separately solved for each hour to obtain market clearing prices ( $\tilde{\Lambda}_t$ ):

$$i \in I, t \in T: \quad \text{Minimize}_{P_{i,t}} \quad \sum_i P_{i,t} \cdot \tilde{\lambda}_{i,t} \quad (1a)$$

$$\text{subject to:} \quad \sum_i P_{i,t} = \mathcal{D}_t \quad (1b)$$

$$0 \leq P_{i,t} \leq \bar{P}_{i,t} \quad (1c)$$

Two datasets corresponding to distinct time periods are used: one for the model calibration (i.e. the training period  $T^*$ ), the other for the actual price simulations (i.e. the test period  $T$ ).

Fig. 1 provides a graphical representation of the simulation method presented in this paper. Internal variables are described later in this section.

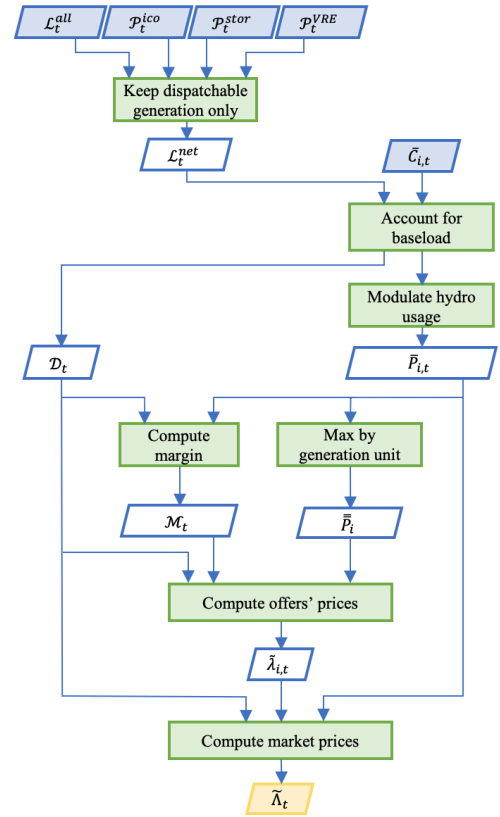


Figure 1. Flowchart of the price simulation methodology

### B. Demand simulation

We first consider the national power consumption covered by the dispatchable generation units of a given country, denoted by  $\mathcal{L}_t^{net}$  and calculated as follows:

$$\mathcal{L}_t^{net} = \mathcal{L}_t^{all} - \mathcal{P}_t^{ico} - \mathcal{P}_t^{stor} - \mathcal{P}_t^{VRE} \quad (2)$$

where  $\mathcal{L}_t^{all}$  is the actual hourly national consumption and  $\mathcal{P}_t^{ico}$ ,  $\mathcal{P}_t^{stor}$ ,  $\mathcal{P}_t^{VRE}$  are respectively the net energy balance of the international interconnections, storage (e.g. hydro pumped storage) and power sources considered non-dispatchable (e.g. solar, wind, hydro run-of-river and poundage).

For the periods and bidding areas where the subtracted contributions may be neglected, this resulting consumption is greater than the volume traded on the day-ahead market, since power transactions also occur over-the-counter and on other organized market places (e.g. futures markets). To account for this, a share of the energy covered by dispatchable generation units that is assumed to correspond to a baseload not traded on the day-ahead market and denoted by  $\mathcal{L}_t^{base}$  is also subtracted. The result is denoted by  $\mathcal{D}_t$  and will be considered as an inelastic demand on the day-ahead market in our model:

$$\mathcal{D}_t = \mathcal{L}_t^{net} - \mathcal{L}_t^{base} \quad (3)$$

An example of baseload calculation is given in the section "Case study".

### C. Supply simulation

For the supply side, we consider all generation units of a given country that have a nominal power above 100 MW and associate them with one offer per hour (sales volumes and prices). We assign to each unit  $i$  the variable  $\bar{P}_i$  that is defined as its maximal available generation capacity observed over the considered period. We group the units according to production type (i.e. nuclear, gas, coal, oil or hydro water reservoir – hereafter referred to as “hydro” and corresponding to the index  $hyd$ ).

#### 1) Volumes

The offer’s volume is set to the residual available generation capacity of generation units ( $\bar{P}_{i,t}$ ). This capacity is equal to the difference between the actual available generation capacity ( $\bar{C}_{i,t}$ ) and the power devoted to the baseload (which might be null depending on the type considered and the assumptions of the modeler).

For hydro, the residual available generation capacity is adapted to reflect the management constraints related to the limited stock of water (i.e. a restricted amount of available energy). In carrying out the transformation (4), we account for the strategic use of this energy reserve based on market clearing prices.

$$\bar{P}_{i,t} = \mathcal{P}^{*,hyd} \cdot \frac{(\alpha \cdot \tilde{\Lambda}_t^* + \beta)}{\sum_t (\alpha \cdot \tilde{\Lambda}_t^* + \beta)} \cdot \frac{\bar{P}_i}{\sum_i \bar{P}_i}, i \in I^{hyd}, t \in T \quad (4)$$

where  $\alpha$  and  $\beta$  are the parameters of the linear regression computed over the training period, using the observed market prices ( $\Lambda_t^*$ ) and the aggregated hydro production at the national level ( $\mathcal{P}_t^{*,hyd}$ ) at time  $t$ :

$$\mathcal{P}_t^{*,hyd} = \sum_i P_{i,t}, \quad i \in I^{hyd}, t \in T^* \quad (5a)$$

$$\mathcal{P}_t^{*,hyd} = \alpha \cdot \Lambda_t^* + \beta, \quad t \in T^* \quad (5b)$$

$$\mathcal{P}^{*,hyd} = \sum_t \mathcal{P}_t^{*,hyd}, \quad t \in T^* \quad (5c)$$

#### 2) Prices per unit of energy

For each hour, a global reserve margin denoted by  $\mathcal{M}_t$  is computed; this margin is defined as the difference between the sum of all residual available generation capacities  $\bar{P}_{i,t}$  at time  $t$  and the residual demand  $\mathcal{D}_t$ .

We also associate the variable  $\bar{S}_i$  with each generation unit. To do so, we first split the units according to production type, then sort the units by ascending creation date, and finally, define  $\bar{S}_i$  as the cumulative sum of  $\bar{P}_i$ .

All prices are then obtained by first applying the following equation for each generation unit:

$$\tilde{\lambda}_{i,t} = a_t \cdot \bar{S}_i + b_t \cdot \mathcal{M}_t + c_t \quad (6)$$

where the stack parameters ( $a_t, b_t, c_t$ ) have been statistically calibrated for each production type as follows. For each hour in the training period, having computed the supply offers and in the knowledge of  $\mathcal{D}_t$ , we determine the type of the most

expensive offer accepted according to our model (herein called the marginal type). For every type, and considering only the hours where the type is marginal, we identify the parameters using the training data and (6). This process is iteratively performed using the new set of parameters to once again simulate prices over the training period. We select the set of parameters leading to the best performance (i.e. lowest root mean square error of market price simulations) over the whole training period.

From the selected hours used to calibrate the stack parameters, we also retain the minimal and maximal observed market prices. As a final tuning of the offers, we cap the simulated prices obtained by (6) for the test period to the minimal and maximal observed prices for each production type.

### D. Market clearing prices simulation

Market clearing prices are obtained by first solving (1), then we adjust them by subtracting a corrective term resulting from the analysis of residuals (i.e. for each hour of the day and day of the week a linear regression between the observed and simulated prices over the modeling period is performed to be used as corrective term, the regression parameters are then reused on the testing period).

## IV. CASE STUDY

### A. Materials

A study of the French bidding zone from 2013-2015 was conducted using market data provided by EPEX SPOT and power system data provided by RTE.

### B. Data processing

The simulations were performed using R programming language [8] and the RStudio integrated development environment [9]. The approach exposed in the previous section was applied. Training and test durations of a calendar year were considered.

In this case study: the contribution of biomass has been neglected, all storage capacities are hydro pumped storage, while hydro run-of-river and poundage are considered to be non-dispatchable.

We assume that only nuclear is used as a baseload. We want to define a baseload that has a linear evolution within a one-week period. So, for each week, we consider a constraint on the demand (i.e. the minimal of  $\mathcal{L}_t^{net}$  resulting from (2) that we multiply by a coefficient denoted by  $k^{nuc}$  corresponding to a maximal share of consumption covered by nuclear power) and a constraint on the supply (i.e. the minimal aggregated power available from all nuclear generation units). To define the baseload, we maintain the minimum value between the demand and supply constraints for each week and then linearly interpolate between the weeks in order to generate the full time series over the whole year as illustrated in Fig. 2.

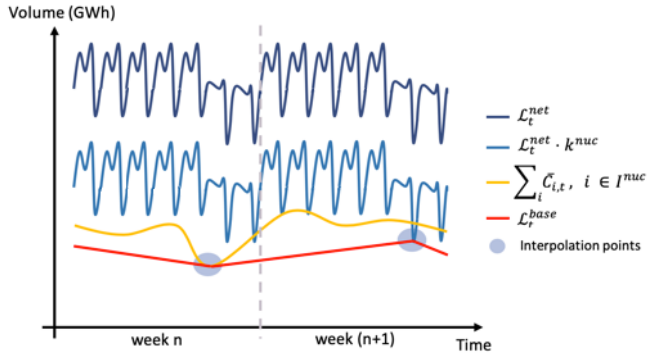


Figure 2. Estimation of the nuclear baseload

Once we have computed the nuclear baseload, we have to individually define a residual available generation capacity for each nuclear generation unit. To do so, we consider that each unit contributes to the baseload proportionally to its  $\bar{P}_i$  and then reduce this value and the available capacity accordingly.

### C. Results and discussion

#### 1) Global accuracy assessment

The performance over the whole period is estimated thanks to the root mean square error and the mean absolute error. The results are summarized in Table 1. The performance of the model over its own training period is given as an indication.

TABLE I. PERFORMANCE METRICS

Metric	Training year	Test year		
		2013	2014	2015
RMSE <sup>a</sup> (€/MWh)	2013	9.7	9.3	9.4
	2014	12.0	6.7	7.2
	2015	11.6	7.7	6.2
MAE <sup>b</sup> (€/MWh)	2013	6.6	7.4	7.2
	2014	8.3	5.1	5.7
	2015	8.0	6.0	4.8

a. Root Mean Square Error  
b. Mean Absolute Error

The time series relative to the best and worst performances in terms of RMSE are presented below. In both cases, this corresponds to the training year 2014, and respectively 2015 and 2013 for the test years.

#### 2) Time series

As shown in Fig. 3, the model reproduced accurate values over the year 2015: a mean of €38.5/MWh instead of €37.2/MWh and a standard deviation of €13.0/MWh for both the simulation and the real data. Seasonal trends can be observed.

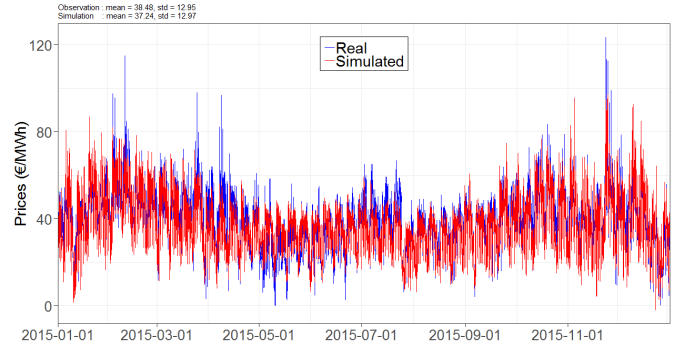


Figure 3. Day-ahead electricity prices in France in 2015

The focus on February 2015 shown in Fig 4. reveals that the daily patterns are reproduced and highlights that the price spikes are not accurately reproduced.

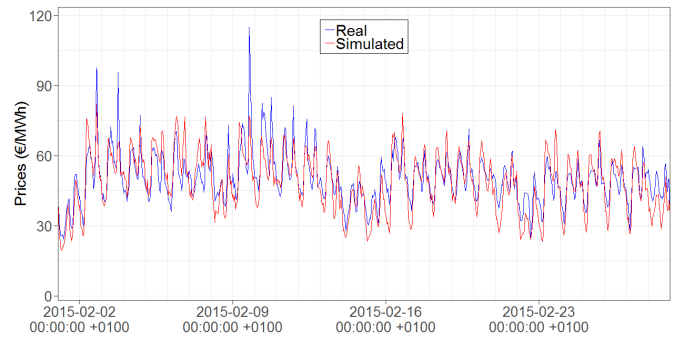


Figure 4. Day-ahead electricity prices in France in February 2015

Fig. 5 helps us to understand the limits of the model. The simulations feature roughly the same means in 2015 and 2013 (respectively €37.2/MWh and €37.9/MWh) while the difference between the real prices is more significant (respectively €38.5/MWh and €43.2/MWh). In addition, the negative and positive price spikes are not well reproduced. Besides these limitations, the global trends appear to be accurately reflected by the simulated prices.

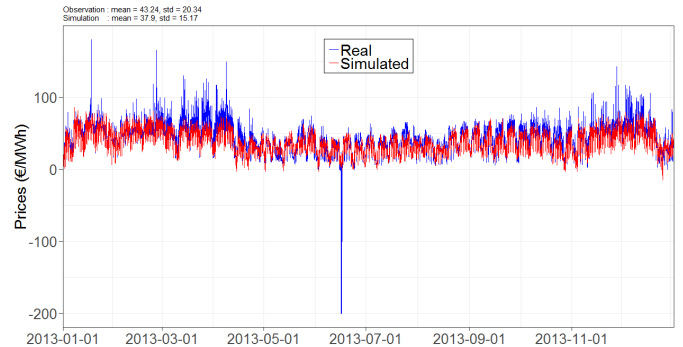


Figure 5. Day-ahead electricity prices in France in 2013

Fig. 6 confirms that even for February 2013, the daily patterns are captured by the model.

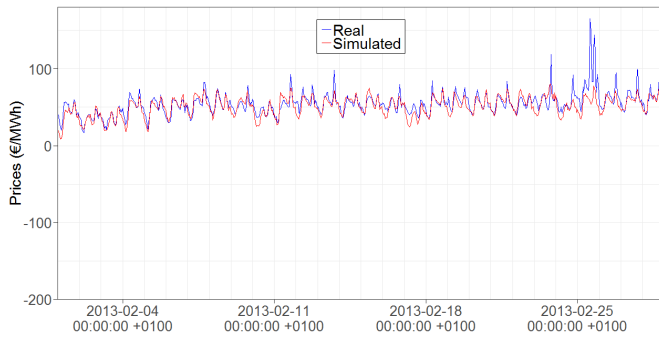


Figure 6. Day-ahead electricity prices in France in February 2013

### 3) Critical analysis

The model highlights the link between the states of the power system and the market prices. The few input variables and the linear modeling technique are sufficient to reveal interesting price patterns. It is to be noted that the case study focuses on a period of three years in one country; a more global study is in preparation to reveal patterns consistent across years and areas. Economic variables reflecting the evolution of fuel and emission costs, in addition to phenomena such as inflation, will be required to account for the evolution of electricity supply costs over a longer time span.

To better simulate price spikes, the current method could be supplemented by a non-linear modeling approach and/or additional inputs, for example related to weather conditions.

## V. CONCLUSION AND PERSPECTIVES

We proposed a method to statistically calibrate a structural model that generates market prices with an hourly resolution over long time periods. Considering relatively few variables related to the power system and using only the observed market clearing prices as financial information, we simulated time series that display interesting characteristics for the French bidding area between 2013 and 2015. Thanks to the European regulation [10], the input variables used by this model are made available by ENTSO-E for all EU countries, implying that this method could be applied to other markets.

The main limitation of the model is related to price spikes, which are not properly simulated. This limitation might be resolved by adding regime-switching features to the model in specific conditions or by adding non-linearities. Moreover, a more detailed consideration of the management of storage

and interconnections should increase accuracy. Accounting for the effect of unusual events such as atypical weather conditions or sudden power plant outage will also be part of future extensions.

In order to use the model for prospective applications on longer time horizons, the evolution of costs should be integrated to the model, as well as the effect of the ever-increasing share of intermittent renewable energy resources. Forecasts could then be generated using simulated input data based on expected scenarios. Such hourly forecasts are crucial to assess expected future revenues and thus evaluate the viability of projects and policies.

## ACKNOWLEDGMENT

We would like to thank RTE and EPEX SPOT for the provision of data.

## REFERENCES

- [1] "ACER guidance on the application of REMIT and transaction reporting," Issue No. 15 (updated version), Q4 2018.
- [2] R. Carmona, M. Coulon, and D. Schwarz, "Electricity price modeling and asset valuation: a multi-fuel structural approach," *Math. Finance Econ.*, vol. 7, no. 2, pp. 167–202, Mar. 2013.
- [3] F. Ziel and R. Steinert, "Probabilistic mid- and long-term electricity price forecasting," *Renew. Sustain. Energy Rev.*, vol. 94, pp. 251–266, Oct. 2018.
- [4] C. Ruiz, A. J. Conejo, and D. J. Bertsimas, "Revealing Rival Marginal Offer Prices Via Inverse Optimization," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3056–3064, Aug. 2013.
- [5] R. Aïd, L. Campi, and N. Langrené, "A Structural Risk-Neutral Model for Pricing and Hedging Power Derivatives," *Math. Finance*, vol. 23, no. 3, pp. 387–438, 2013.
- [6] L. Mitridati and P. Pinson, "A Bayesian Inference Approach to Unveil Supply Curves in Electricity Markets," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2610–2620, May 2018.
- [7] R. Chen, I. C. Paschalidis, M. C. Caramanis, and P. Andrianesis, "Learning from Past Bids to Participate Strategically in Day-Ahead Electricity Markets," *IEEE Trans. Smart Grid*, pp. 1–1, 2019.
- [8] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019.
- [9] RStudio Team, *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc., 2018.
- [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*, vol. 163. 2013.