# User Manual of the Low-carbon Expansion Generation Optimization (LEGO) model

S. Wogrin[a,b,*], R. Gaugl[a], T. Klatzer[a], U. Bachhiesl[a], D. Tejada-Arango[c,d]

[a] *Institute of Electricity Economics and Energy Innovation (IEE), Graz University of Technology, Graz, Austria*
[b] *Institute for Research in Technology (IIT), School of Engineering (ICAI), Comillas Pontifical University, Madrid, Spain*
[c] *Energy Management Consulting, AFRY, Madrid, Spain*
[d] *Facultad de Ciencias Económicas y Empresariales (ICADE), Universidad Pontificia Comillas, Madrid, Spain*

---

**Abstract**

This is the user manual of the Low-carbon Expansion Generation Optimization (LEGO) model available on GitHub[1]. LEGO is a mixed integer quadratically constrained optimization problem and has been designed to be a multi-purpose tool, like a Swiss army knife, that can be employed to study many different aspects of the energy sector. Ranging from short-term unit commitment to long-term generation and transmission expansion planning. The underlying modeling philosophies are: modularity and flexibility. Its unique temporal structure allows LEGO to function with either chronological hourly data, or all kinds of representative periods. LEGO is also composed of thematic modules that can be added or removed from the model easily via data options depending on the scope of the study. Those modules include: unit commitment constraints; DC- or AC-OPF formulations; battery degradation; rate of change of frequency inertia constraints; demand-side management; or the hydrogen sector. LEGO also provides a plethora of model outputs (both primal and dual), which is the basis for both technical but also economic analyses. We hereby make LEGO freely available to the scientific community.

---

*Corresponding author:
*URL:* `wogrin@tugraz.at` (S. Wogrin)

[1] https://github.com/IEE-TUGraz/LEGO

Contents


**1 Overview** **4**

**2 Download and installation** **6**

**3 Licenses and software requirements** **6**

**4 LEGO software architecture** **7**

**5 How to run LEGO** **8**

**6 Modularity and Flexibility** **9**

6.1 Modularity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

6.2 Flexibility . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

6.3 Extensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

**7 Input Data** **15**

7.1 General . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

7.1.1 Indices . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

7.1.2 Parameters . . . . . . . . . . . . . . . . . . . . . . . . . . 19

7.2 Electricity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

7.2.1 Demand . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

7.2.2 ThermalGen . . . . . . . . . . . . . . . . . . . . . . . . . . 24

7.2.3 StorageUnits . . . . . . . . . . . . . . . . . . . . . . . . . 28

7.2.4 VREs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

7.2.5 VRE-profiles . . . . . . . . . . . . . . . . . . . . . . . . . 35

7.2.6 FACTS . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

7.2.7 Inflows . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

7.2.8 Weights . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

7.2.9 Hindex . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

7.2.10 DSM-DelayTime . . . . . . . . . . . . . . . . . . . . . . . 41

7.2.11 DSM-Max . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

7.2.12 DSM-shift . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

7.2.13 DSM-shed . . . 44
7.2.14 Network . . . 45
7.2.15 BusInfo . . . 48
7.3 H2-Sector . . . 50
7.3.1 H2_Indices . . . 50
7.3.2 H2_Demand . . . 52
7.3.3 H2_GenUnits . . . 53
7.3.4 H2_Network . . . 55

**8 Outputs** **56**
8.1 General . . . 57
8.1.1 Summary . . . 58
8.1.2 TotalEn . . . 60
8.1.3 GenInvest . . . 60
8.1.4 H2Invest . . . 60
8.1.5 CDSF . . . 61
8.1.6 UC . . . 61
8.1.7 GenP . . . 61
8.1.8 GenQ . . . 62
8.1.9 Charge . . . 62
8.1.10 Curtail . . . 62
8.1.11 StIntra . . . 63
8.1.12 StLevel . . . 63
8.1.13 MC . . . 63
8.1.14 InertDual . . . 63
8.1.15 RoCoF . . . 64
8.1.16 LineP . . . 64
8.1.17 LineQ . . . 64
8.1.18 Volt . . . 65
8.1.19 Angle . . . 65
8.1.20 Profits . . . 65

8.1.21 DSM . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
8.1.22 H2price . . . . . . . . . . . . . . . . . . . . . . . . . . . 66
8.1.23 H2Prod . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
8.1.24 H2Cons . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
8.1.25 H2ns . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

**9 How to construct new cases 67**

**10 APPENDIX A: Nomenclature 71**

**11 APPENDIX B: Mathematical Model Formulation 75**

11.1 Objective function and standard constraints . . . . . . . . . . . . 75
11.2 Demand-side management . . . . . . . . . . . . . . . . . . . . . . 77
11.3 Transmission expansion planning (DC + AC) . . . . . . . . . . . 79
11.3.1 DC-OPF . . . . . . . . . . . . . . . . . . . . . . . . . . 79
11.3.2 AC-OPF . . . . . . . . . . . . . . . . . . . . . . . . . . 80
11.4 Hydrogen . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
11.5 Standard Constraints . . . . . . . . . . . . . . . . . . . . . . . . 84
11.6 Battery degradation . . . . . . . . . . . . . . . . . . . . . . . . 86
11.7 Modeling Inertia . . . . . . . . . . . . . . . . . . . . . . . . . . 87

**12 APPENDIX C: Literature Review 88**

**13 APPENDIX D: Case Studies 92**

13.1 Hourly model versus representative days model . . . . . . . . . . 92
13.2 Impact of demand-side management . . . . . . . . . . . . . . . . 95
13.3 Hydrogen production . . . . . . . . . . . . . . . . . . . . . . . . 98
13.4 Transmission expansion planning DC versus AC . . . . . . . . . . 99

## 1. Overview

To mitigate climate change, we as a society have embarked on the journey towards net-zero energy systems [1, 2]. Ahead of us there lie massive regulatory, social, economic and technical challenges such as the effective coupling of the electric power, heat and gas sectors, as well as the electrification of transport, active demand-side management and many more. Transparent, open-source energy system models (ESMs) are necessary to support the energy transition through more informed and better decision making. For an overview of non-open-source ESMs the reader is referred to [3].

In this paper, we want to introduce a newly developed open-source ESM called Low-carbon Expansion Optimization (LEGO) model. The LEGO model is a mixed-integer quadratically[2] constrained optimization problem (MIQCP) coded in GAMS [4]. The data are read directly from Excel and the results are also written to the same Excel file. LEGO has been designed to be a multi-purpose tool, like a Swiss army knife, that can be employed to study many different aspects of the energy sector. The underlying modeling philosophies are: modularity and flexibility. LEGO also provides a plethora of model outputs (both primal and dual). **You can download and use this model freely, but if you do, please cite [5].**

LEGO is *modular* in the sense that it can be assembled with different thematic blocks that allow for a wide variety of specific analyses. These modules currently include: considering unit commitment (UC) decisions[3]; considering (or not[4]) an electricity network; considering the electricity grid via a DC or an AC optimal power flow (OPF)[5] formulation; considering degradation for battery energy storage systems (BESS) via cycle aging costs; considering rate of change of frequency (RoCoF) system inertia constraints; considering demand-side management (DSM) via load shifting and load shedding; considering the hydrogen sector.

---

[2]It is only quadratic if the SCOP formulation of the AC-OPF is solved, otherwise the model is a mixed-integer linear program (MILP).

[3]This module can be relaxed by solving LEGO as a relaxed Mixed Integer Program (rMIP).

[4]In this case, LEGO is converted into a single-node problem.

[5]In particular, we consider a second-order cone programming (SOCP) approximation of the full AC-OPF.

LEGO's model *flexibility* is currently three-fold: flexibility of running LEGO as a mixed integer program (MIP) and consider discrete decisions (such as lumpy investments and UC) or as a relaxed-MIP, where discrete variables are relaxed; flexibility of running an operation only or an investment (GEP, TEP or GEPTEP)[6] model; flexibility in the representation of time (representative periods, hourly models, time segments/blocks).

The remainder of the user manual is organized as follows. In section 2 we explain how to download and install the LEGO model; section 3 explains what software packages are required and the corresponding licenses; section 4 discusses LEGO's software architecture, and section 5 explains how to run the model. Section 6 discusses some further details about LEGO's modularity and flexibility options. The input data and outputs are listed in sections 7 and 8. Section 9 describes some ways to construct new cases. In the appendix, we first introduce the nomenclature (Appendix A) that is used in the mathematical formulation (Appendix B) of the LEGO model. This is followed by a literature review of ESMs (Appendix C) and some sample case studies (Appendix D).

## 2. Download and installation

1. If you haven't done so, download GAMS to your computer from[7]
2. Follow this link: https://github.com/IEE-TUGraz/LEGO
3. Download the ZIP file using the 'Code' button
4. Unzip and keep all files in the same folder

## 3. Licenses and software requirements

In order to run LEGO, the user requires the following software:

[6]We consider candidate investments in both generation expansion planning (GEP) and transmission expansion planning (TEP)

[7]https://www.gams.com/download/

- Microsoft Excel. This is where we store model input data, and to where we output model results.
- GAMS (General Algebraic Modeling System). The optimization model LEGO is coded in GAMS. GAMS grants free (but limited) academic licenses. Depending on the license the problem size that can be solved is also limited (e.g. up to 2000 variables and 2000 equations). To that purpose, we have included a test case using 1 representative day (LEGO-Base-Case-Study-1LRP.zip), which runs without requiring the professional GAMS license. For the 7 representative day and the hourly case, a full GAMS license will be necessary.
- Numerical solvers: Within GAMS we use the numerical solver CPLEX in order to solve Mixed Integer Quadratically Constrained (MIQCP) problems. Note that Gurobi or any other MIQCP solver could also be used for LEGO. Hence, a CPLEX license is necessary to solve large-scale instances of LEGO.

We denote LEGO as *open-source* model because the GAMS source code is freely available. Independently of the case study data applied, the source code remains the same. Moreover, small numerical cases (like the 1LRP case) run with the free academic license. Only solving large-scale numerical examples requires a professional GAMS & CPLEX license. For academic institutions these licenses even come at a reduced rate.

## 4. LEGO software architecture

The LEGO code runs entirely in GAMS (using MIQCP solvers such as CPLEX or Gurobi). Data and results are read from / exported to Microsoft Excel via gdx as indicated in Figure 2. For a more detailed description of the input data and outputs, the reader is referred to sections 7 and 8.

Within the LEGO.gms file, we follow a particular nomenclature convention. All data that is read from Excel (ranges) is imported into parameter tables whose

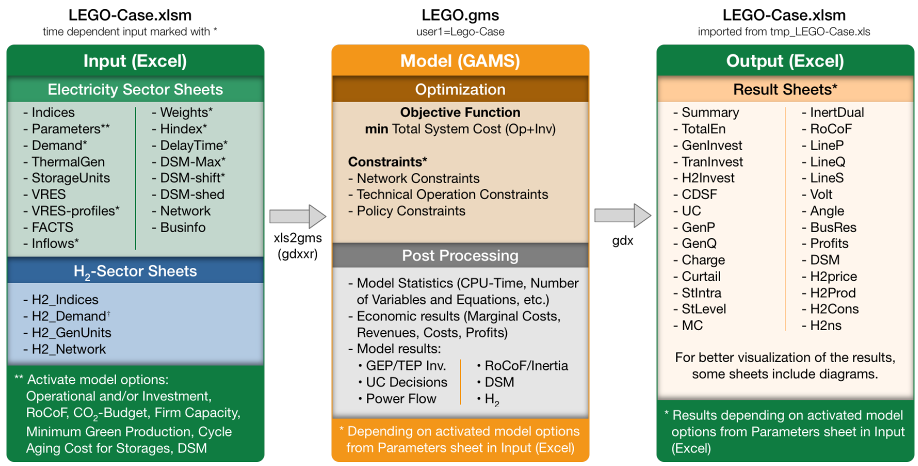


Figure 1: LEGO software architecture.

name starts with "t". Here the data can be scaled etc. These tables, however, are only for data import from Excel and are not used in the optimization model per se. The parameters used in the optimization model all start with a "p". All variables have names that start with "v", and all constraints (denominated equations in GAMS) have names that start with an "e". After solving the optimization problem, some ex-post calculations (of interest) are carried out and results of such calculations (such as profits) are written into output parameters (whose names also start with "p").

## 5. How to run LEGO

1. Unzip the data files (e.g. LEGO-Base-Case-Study-1LRP.zip) into .xlsm
2. Double click on LEGO.gms
3. In GAMS, load your data file by writing 'user1=LEGO-Base-Case-Study-1LRP' in the GAMS parameters box in the upper right hand corner of the IDE or GAMS Studio screen (without the file extension)
4. Run the model with F9, or by clicking on the run button
5. Model outputs are automatically written into temporary Excel files (e.g. tmp_LEGO-Base-Case-Study-1LRP.xlsx), which can be imported into the

data file (e.g. LEGO-Base-Case-Study-1LRP.xlsm) by clicking the 'Load' button on the Menu sheet

Note: Avoid using spaces or special characters in your file names

## 6. Modularity and Flexibility

This section contains further details about how to utilize LEGO's modularity and flexibility via data, and finally, discusses some possible room for extensions.

### *6.1. Modularity*

In the introduction, we have already discussed the different thematic modeling blocks[8] that LEGO offers. Each of these modules can be combined freely with the others, and it is as simple as activating a (yes/no) option in the data file in Excel. In the LEGO data file, there is a sheet named Parameters, shown in Figure 2, where these options can be chosen. For example, one can carry out a DC-OPF and study the impact on the production of hydrogen when allowing DSM.

Let us briefly repeat the nomenclature for types of mathematical optimization problems used here: linear program (LP) - the simplest type of model where all variables are continuous and all constraints are linear. LPs are the easiest types of models to solve and (for the same amount of variables) have the fastest CPU times. Quadratically constrained problems (QCP) - all variables are continuous but constraints can be quadratic. If constraints are convex, then this is also a relatively easy type of problem that can be solved numerically very efficiently. Mixed integer program (MIP) - a linear model where some variables can be discrete and some continuous. In general, discrete variables complicate CPU times by orders of magnitude (with respect to LPs and QCPs), because corresponding solution algorithms (such as branch and bound) can require a lot of memory.

[8] UC, DC/AC-OPF, RoCoF, battery degradation, the hydrogen sector, DSM, etc.

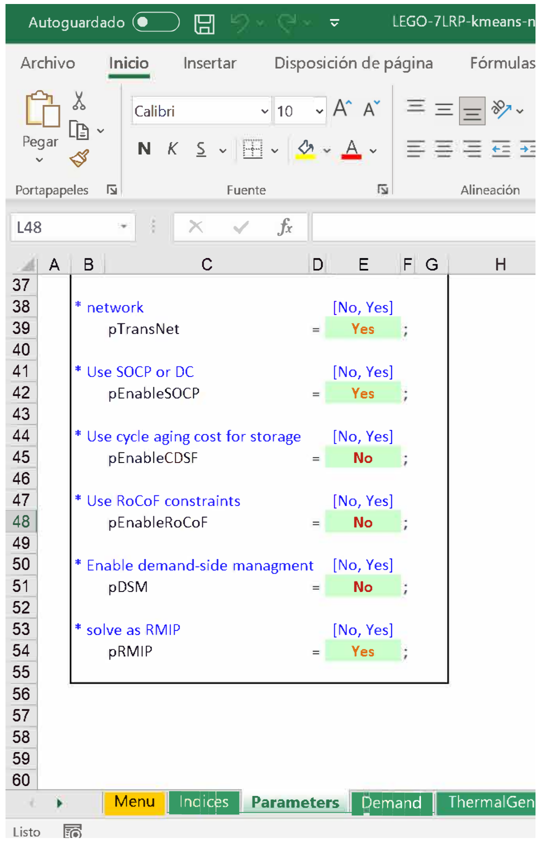


Figure 2: LEGO parameter sheet with model options.

In Table 1 we present the different existing modules in LEGO, and the corresponding type of mathematical formulation. With respect to combining these modules, in principle, all combinations should be possible. For example, combining a MIP formulation of the unit commitment with the AC-OPF (QCP formulation) yields an MIQCP type of problem. Some modules can be run as a MIP or a relaxed rMIP. For example, unit commitment in general has to be formulated as a discrete problem, i.e. a MIP. However, sometimes it might make sense - for example due to numerical tractability - to relax the UC formulation as this yields an LP model. A relaxed UC might be able to give a rough lower bound of how many thermal plants to start up at a particular hour, without having to endure long CPU times waiting for the full MIP version to finish.

With respect to GEP (or TEP with DC-OPF), an rMIP approximation might give an indication of how many MW of new investments might be necessary. However, the TEP with AC-OPF can only be solved using a MIP approach. An rMIP formulation is not meaningful here.

Let us briefly discuss the relation between the CPU time and the type of problem. In general, the relation is $CPU(LP) < CPU(QCP) < CPU(MIP) < CPU(MIQCP)$, however, it is difficult to make an a-priori estimation of which set of constraints will complicate CPU time more, e.g. UC or GEP for example. First, that depends on the amount of discrete variables (how many thermal power plants are there versus how many candidate investments are there). Second, even if both formulations incurred the same amount of discrete variables, it is hard to compare MIPs. It very much depends on input data. For example, two models with the exact combination of modules can take either 5 minutes or several hours depending on, for example, whether the parameter for green production is 40 or 100%. Hence, the following statements are some general observations we have made, but they do not necessarily have to apply to all problem instances. Including RoCoF constraints have the highest computational burden. Then, the AC-OPF when combined with other MIP features. And UC usually has a worse impact than GEP.

### *6.2. Flexibility*

Model flexibility, just like LEGO's modularity, can be tapped into easily via the data. This means that LEGO users do not have to bother to change the code or the model itself. They can simply access all of LEGO's flexibility via data input. Let us briefly describe how this can be achieved.

LEGO has been designed as a Mixed Integer Problem (MIP); the integrality stems from planning variables (i.e., investments) and operational (i.e., unit commitment, UC) decisions. Relaxing integrality on these two sets of variables renders a relaxed-MIP (rMIP) framework that still has physical meaning. This can be obtained using the *pRMIP* option shown in Figure 2.

With respect to running LEGO as an operation only or an investment model,

| **Module** | **Problem type** | **Activation (sheet, parameter)** |
|---|---|---|
| Network<br>DC-OPF/AC-OPF | LP/QCP | Parameters, pTransNet (Y/N)<br>Parameters, pEnableSOCP (Y=AC/N=DC) |
| RoCoF | MIP | Parameters, pEnableRoCoF (Y/N) |
| BESS degradation | LP | Parameters, pEnableCDSF (Y/N) |
| DSM | LP | Parameters, pDSM (Y/N) |
| Hydrogen | LP | Parameters, pEnableH2 (Y/N) |
| UC | MIP | Parameters, pEnableH2 (Y/N) |
| GEP | MIP | Thermalgen, EnableInvest (0,1)<br>StorageUnits, EnableInvest (0,1), MaxInvest [0-N]<br>VRES, EnableInvest (0,1), MaxInvest [0-N] |
| TEP | MIP | Network, FixedCost>0 (is candidate line) |
| Relax model | rMIP or MIP | Parameters, pRMIP (Y/N) |

Table 1: LEGO modules, type of mathematical formulation and how/where to activate.

in the corresponding data sheet as seen in Figure 3, candidate transmission lines can be explicitly specified. If there are none, LEGO does not consider TEP. As for GEP, in LEGO one can specify how many of the units are existing (via parameter *ExisUnits*), whether or not generation expansion is possible for this type of unit (via parameter *EnableInvest*), and what is the maximum amount of this type of unit that can be built (via parameter *MaxInvest*), all of which can be seen in Figure 4.

Let us now discuss in detail the type of temporal flexibility that LEGO allows for. In many energy modeling applications, such as running a UC model for example, the desired temporal representation are chronological hours. Especially in short-term applications. Medium, or long-term models, however, might resort to representative periods (days or weeks), or even time slices [6] or load periods [7] to represent very long time horizons that would lead to computationally intractable models if chronological hourly periods were to be considered. Depending on the data input, LEGO allows for all these types of studies without

* Network Information

| From bus i | To bus j | Circuit ID | InService [0 - 1] | R [p.u.] | X [p.u.] | Bc [p.u.] | TapAngle [º] | TapRatio [p.u.] | Pmax [MW] | FixedCost [M€] |
|---|---|---|---|---|---|---|---|---|---|---|
| Node_1 | Node_6 | c1 | 1 | 0.0020 | 0.0200 | 0.0360 | 0.0000 | 1.0000 | 800 | |
| Node_2 | Node_3 | c1 | 1 | 0.0040 | 0.0300 | 0.0540 | 0.0000 | 1.0000 | 800 | |
| Node_2 | Node_6 | c1 | 1 | 0.0060 | 0.0400 | 0.0720 | 0.0000 | 1.0000 | 800 | |
| Node_3 | Node_4 | c1 | 1 | 0.0080 | 0.0500 | 0.0900 | 0.0000 | 1.0000 | 800 | |
| Node_3 | Node_6 | c1 | 1 | 0.0070 | 0.0600 | 0.1080 | 0.0000 | 1.0000 | 800 | |
| Node_4 | Node_5 | c1 | 1 | 0.0050 | 0.0700 | 0.1260 | 0.0000 | 1.0000 | 800 | |
| Node_4 | Node_6 | c1 | 1 | 0.0030 | 0.0800 | 0.1440 | 0.0000 | 1.0000 | 800 | |
| Node_4 | Node_9 | c1 | 1 | 0.0010 | 0.0700 | 0.1260 | 0.0000 | 1.0000 | 800 | |
| Node_6 | Node_7 | c1 | 1 | 0.0030 | 0.0600 | 0.1080 | 0.0000 | 1.0000 | 800 | |
| Node_6 | Node_8 | c1 | 1 | 0.0050 | 0.0500 | 0.0900 | 0.0000 | 1.0000 | 800 | |
| Node_7 | Node_8 | c1 | 1 | 0.0070 | 0.0400 | 0.0720 | 0.0000 | 1.0000 | 800 | |
| Node_8 | Node_9 | c1 | 1 | 0.0060 | 0.0300 | 0.0540 | 0.0000 | 1.0000 | 800 | |
| Node_1 | Node_4 | c1 | 1 | 0.0060 | 0.0300 | 0.0540 | 0.0000 | 1.0000 | 800 | |
| * candidate lines (any line with a fixed cost associated) | | | | | | | | | | |
| Node_4 | Node_5 | c2 | 1 | 0.0050 | 0.0700 | 0.1260 | 0.0000 | 1.0000 | 800 | 100 |

VRES | VRES-profiles | FACTS | Inflows | Weights | Hindex | DSMindex | DSM-profiles | DSM-shift | DSM-shed | Network | BusInfo | Summary | TotalEn | GenInv ...

Figure 3: LEGO power network data including existing and candidate lines.

* variable renewable energy sources

| | ExisUnits [0-N] | MaxProd [MW] | EnableInvest [0,1] | MaxInvest [0-N] | InvestCost [$/MW/year] | OMVarCost [$/MWh] | FirmCapCoef [p.u.] | Qmax [MVAr] | Qmin [MVAr] | InertiaConst [s] |
|---|---|---|---|---|---|---|---|---|---|---|
| Wind | 0 | 100 | 1 | 290 | 72641.8 | 2.0 | 0.07 | | | 0 |
| Wind2 | 0 | 100 | 1 | 290 | 80000.0 | 5.0 | 0.07 | | | 2 |
| Solar | 0 | 100 | 1 | 300 | 84467.2 | 0.0 | 0.14 | | | 0 |
| Solar2 | 0 | 100 | 1 | 300 | 84467.2 | 0.0 | 0.14 | | | 0 |

Figure 4: LEGO variable renewable energy sources data.

having to adapt the model code and formulation, simply via the data.

In order to achieve this, we define three different temporal indices: $p, k, rp$, which are specified in the data sheet Indices, as seen in Figure 5, and which we will use throughout the model formulation. Index $p$ represents the actual chronological periods (which are usually hourly but could even have a shorter duration); $rp$ are the representative periods used; and finally, index $k$ corresponds to the chronological periods within the representative period $rp$. In order for those indices, and corresponding variables, to have proper meaning, we also introduce weight parameters $W_{rp}^{RP}$ and $W_k^K$, which represent the weight of this representative period and the weight of the period $k$ within each $rp$ respectively. Those weights are specified in data sheet Weights. Finally, there is also a mapping $\Gamma(p, rp, k)$ that relates each actual period $p$ to its representative period $rp$ and period $k$, which is specified in data sheet Hindex.

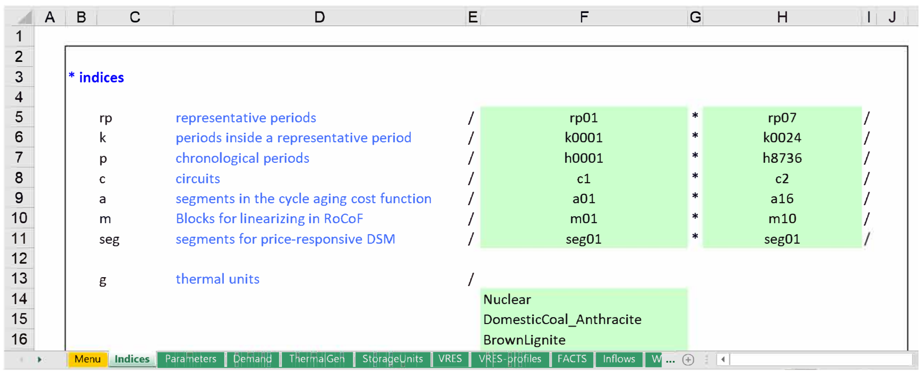


Figure 5: LEGO index and dynamic sets data.

In the example shown in Figure 5, we are representing 8736 chronological hours ($p$) by using 7 representative periods ($rp$), each of which has a duration of 24 hours ($k$). Hence, the corresponding data file uses 7 representative days to capture one year of 8736 chronological hours. If instead, we wanted to run LEGO for the exact hourly model, we would need a data file where index $rp$ is 1 (one representative period which is the entire year). Index $k$ are the chronological periods within the year, so $k$ ranges from 1 to 8736, which would coincide with $p$ in this case. All the weights, both $W_{rp}^{RP}$ and $W_k^K$, are equal to 1 and $\Gamma(p, rp, k)$ in this case simply associates $p$ with the corresponding $k$. Note that both data files could correspond to the same original time series of data, e.g., hourly demand data, hourly wind or solar profiles, or hydro inflow data. Given these original time series, one can either use the hourly data directly yielding a LEGO data file corresponding to an exact hourly model, or, one could apply a clustering procedure to the time series and use the clustered data of representative days for example, in a LEGO data file with representative periods.

### *6.3. Extensions*

LEGO is a powerful tool, but every model can be extended. Some obvious extensions would be making LEGO stochastic and adding new modules.

Stochasticity: currently, LEGO does represent some type of short-term uncertainty by representing renewable profiles (which do cover some intermittency), but LEGO is not a stochastic model per se. Introducing stochasticity is labor-intense because of several reasons. First, a new index (i.e. scenario) would have to be added to almost every variable and parameter of the existing model. We believe this could be done relatively fast, however, it would also entail to extend data sheets to depend on scenarios (and making sure Excel does not crash because of large amounts of data let alone finding meaningful data), and more importantly, since corresponding optimization models would become even larger, a straight forward call to a solver would probably no longer be sufficient, and decomposition techniques would have to be applied. In conclusion, this is one of the extensions that are in the pipeline but it is not trivial.

Adding new modules: Since LEGO has been designed as a modular software, adding new sectors (e.g. gas/hydrogen/heat) is quite straight forward. The majority of the formulation of these sectors might even be independent of the existing detailed formulation of the power system. However, the parts where sectors interconnect, need to be handled with care. Meaning that, to program the interconnection, the user will have to fully understand the existing code (in order to identify, which existing constraints need to be altered and how). For example, when we added hydrogen to the existing power-system-only model, we needed to add the power consumption of electrolyzers in the existing power balance. In our opinion, however, the effort to achieve this is relatively low.

## 7. Input Data

This section provides a list with all the names of the sheets that contain inputs; and for each sheet the input data with a short description, unit and corresponding model parameter (in GAMS) is described. Note that the name of the model parameter used in GAMS does not coincide with the corresponding name of the mathematical formulation given in the appendix for reasons of simplicity and legibility. The open-source data sets are based on the StarNet Lite demo version for long-term planning developed by Prof. Andres Ramos at IIT-Comillas (https://pascua.iit.comillas.edu/aramos/starnet.htm), but that data can be changed depending on the users system data.

### *7.1. General*

This section contains the description of the Indices and Parameters sheets. Note that the title of each subsection coincides with the name of the corresponding data sheet in the input Excel.

*7.1.1. Indices*

Definition of indices.

| | |
|---|---|
| | **rp** |
| **Description** | Number of representative periods |
| **Set** | `rp` |
| | **k** |
| **Description** | Number of chronological periods within a representative period |
| **Set** | `k` |
| | **p** |
| **Description** | Chronological periods |
| **Set** | `p` |
| | **c** |
| **Description** | Maximum number of circuits in network |
| **Set** | `c` |
| | **a** |
| **Description** | Segments in the cycle aging cost function used for approximating degradation |
| **Set** | `a` |
| | **m** |
| **Description** | Blocks used in linearization for RoCoF formulation |
| **Set** | `m` |
| | **seg** |
| **Description** | Segments for price-responsive DSM |
| **Set** | `seg` |
| | **sec** |

| | |
|---|---|
| **Description** | Sectors for DSM shifting |
| **Set** | `sec` |
| | **g** |
| **Description** | Initialization of generators divided into thermal units, FACTS, storage units and variable renewable energy sources |
| **Set** | `g` |
| | **i** |
| **Description** | Nodes/buses of electric power system (j is the alias of i) |
| **Set** | `i` |
| | **is** |
| **Description** | Definition of slack node |
| **Set** | `is` |
| | **gi(g,i)** |
| **Description** | Definition of which generator g is connected to which node i |
| **Set** | `gi(g,i)` |
| | **tec** |
| **Description** | Index of generation technologies in power sector |
| **Set** | `tec` |
| | **gtec(g,tec)** |
| **Description** | Relation among generator g and technology tec |
| **Set** | `gtec(g,tec)` |

*7.1.2. Parameters*

Sheet that contains the definition of certain parameters and available options.

| **cost of energy not served** | |
|---|---|
| **Description** | Cost/Penalty of energy not served |
| **Unit** | €/MWh |
| **Parameter** | `pENSCost` |
| **cost of hydrogen not served** | |
| **Description** | Cost/Penalty of hydrogen not served |
| **Unit** | €/kg |
| **Parameter** | `pH2NSCost` |
| **Budget for $CO_2$** | |
| **Description** | Maximum $CO_2$ budget for electricity production per year |
| **Unit** | Mt $CO_2$/y |
| **Parameter** | `pCO2Budget` |
| **cost of $CO_2$ emissions** | |
| **Description** | Cost of $CO_2$ emissions |
| **Unit** | €/t $CO_2$ |
| **Parameter** | `pCO2Price` |
| **cost of $CO_2$ penalty** | |
| **Description** | Cost of $CO_2$ penalty when exceeding $CO_2$-Budget |
| **Unit** | €/t $CO_2$ |
| **Parameter** | `pCO2Penalty` |
| **moving window for LRP** | |
| **Description** | Moving window (for linked representative days) indicating after how many hours the inter-period storage constraints are imposed |

| | |
|---|---|
| **Unit** | h |
| **Parameter** | `pMovWind` |

| **base power** | |
|---|---|
| **Description** | Base power |
| **Unit** | MVA |
| **Parameter** | `pSBase` |

| **slack bus reference voltage** | |
|---|---|
| **Description** | Slack bus reference voltage |
| **Unit** | p.u. |
| **Parameter** | `pSlackVolt` |

| **reserve up** | |
|---|---|
| **Description** | Percentage of total system demand that needs to be allocated for providing upward reserve |
| **Unit** | % |
| **Parameter** | `p2ndResUp` |

| **reserve down** | |
|---|---|
| **Description** | Percentage of total system demand that needs to be allocated for providing downward reserve |
| **Unit** | % |
| **Parameter** | `p2ndResDw` |

| **minimum required inertia** | |
|---|---|
| **Description** | Minimum of required inertia |
| **Unit** | s |
| **Parameter** | `pMinInertia` |

| **base frequency** | |
|---|---|
| **Description** | Base Frequency |
| **Unit** | Hz |

| | |
|---|---|
| **Parameter** | `pBaseFreq` |

| **minimum green production** | |
|---|---|
| **Description** | Minimum percentage of green electricity production |
| **Unit** | % |
| **Parameter** | `pMinGreenProd` |

| **minimum firm capacity** | |
|---|---|
| **Description** | Minimum percentage of firm capacity |
| **Unit** | % |
| **Parameter** | `pMinFirmCap` |

| **maximum angle difference** | |
|---|---|
| **Description** | Maximum angle difference |
| **Unit** | ° |
| **Parameter** | `pMaxAngleDiff` |

| **network** | |
|---|---|
| **Description** | Consider network or calculate single node system |
| **Unit** | Boolean |
| **Parameter** | `pTransNet` |

| **SOCP versus DC** | |
|---|---|
| **Description** | Consider SOCP (simplified AC) or DC optimal power flow calculation |
| **Unit** | Boolean |
| **Parameter** | `pEnableSOCP` |

| **cycle aging cost for storage** | |
|---|---|
| **Description** | Consider cycle aging cost for storage or not |
| **Unit** | Boolean |
| **Parameter** | `pEnableCDSF` |

| RoCoF | |
|---|---|
| **Description** | Consider use Rate of Change of Frequency constraints |
| **Unit** | Boolean |
| **Parameter** | `pEnableRoCoF` |

| consider H2 | |
|---|---|
| **Description** | Consider $H_2$ sector |
| **Unit** | Boolean |
| **Parameter** | `pEnableH2` |

| consider $CO_2$ | |
|---|---|
| **Description** | Consider $CO_2$ constraints |
| **Unit** | Boolean |
| **Parameter** | `pEnableCO2` |

| Enable demand-side management | |
|---|---|
| **Description** | Enable demand-side management |
| **Unit** | Boolean |
| **Parameter** | `pDSM` |

| solve as rMIP | |
|---|---|
| **Description** | Solve LEGO as relaxed mixed integer problem (if Yes: Commitment of thermal power plants can be between 1 and 0; if No: Commitment of thermal power plants if either 1 or 0) |
| **Unit** | Boolean |
| **Parameter** | `pRMIP` |

### *7.2. Electricity*

The sheets that contain input data for the electricity sector are explained in this section.

#### *7.2.1. Demand*

Hourly power demand for each node and representative period.

| **Demand** | |
|---|---|
| **Description** | Hourly power demand per node |
| **Unit** | MW |
| **Parameter** | `pDemandP(rp,k,i)` |

*7.2.2. ThermalGen*

Thermal generators with their properties.

| ExisUnits | |
|---|---|
| **Description** | Specifies the unit as existing [1] or not [0] |
| **Unit** | Boolean |
| **Parameter** | `pExisUnits(t)` |

| MaxProd | |
|---|---|
| **Description** | Maximum active power output of the thermal generator |
| **Unit** | MW |
| **Parameter** | `pMaxProd(t)` |

| MinProd | |
|---|---|
| **Description** | Minimum active power output of the thermal generator |
| **Unit** | MW |
| **Parameter** | `pMinProd(t)` |

| RampUp | |
|---|---|
| **Description** | Rate at which the thermal generator can increase its power output per time step |
| **Unit** | MW |
| **Parameter** | `pRampUp(t)` |

| RampDw | |
|---|---|
| **Description** | Rate at which the thermal generator can decrease its power output per time step |
| **Unit** | MW |

| | |
|---|---|
| **Parameter** | `pRampDw(t)` |

| **Qmax** | |
|---|---|
| **Description** | Maximum reactive power output of the thermal generator |
| **Unit** | MVAr |
| **Parameter** | `pMaxGenQ(t)` |

| **Qmin** | |
|---|---|
| **Description** | Minimum reactive power output of the thermal generator |
| **Unit** | MVAr |
| **Parameter** | `pMinGenQ(t)` |

| **InertiaConst** | |
|---|---|
| **Description** | Inertia constant H of the thermal generator |
| **Unit** | s |
| **Parameter** | `pInertiaConst(t)` |

| **FuelCost** | |
|---|---|
| **Description** | Fuel cost of the thermal generator. Used to calculate `pSlopeVarCost(t)`, `pInterVarCost(t)` and `pStartupCost(t)` |
| **Unit** | €/Mcal |
| **Parameter** | `tThermalGen(t,'FuelCost')` |

| **SlopeVarCost** | |
|---|---|
| **Description** | Slope variable cost of the thermal generator |
| **Unit** | Mcal/MWh |
| **Parameter** | `tThermalGen(t,'SlopeVarCost')` |

| InterVarCost | |
|---|---|
| **Description** | Intercept variable cost of the thermal generator |
| **Unit** | Mcal/h |
| **Parameter** | `tThermalGen(t,'InterVarCost')` |

| OMVarCost | |
|---|---|
| **Description** | Operation and maintenance cost of the thermal generator |
| **Unit** | €/MWh |
| **Parameter** | `pOMVarCost(t)` |

| StartupCost | |
|---|---|
| **Description** | Start up cost of the thermal generator |
| **Unit** | Mcal |
| **Parameter** | `tThermalGen(t,'StartupCost')` |

| EFOR | |
|---|---|
| **Description** | Equivalent forced outage rate of the thermal generator |
| **Unit** | p.u. |
| **Parameter** | `pEFOR(t)` |

| EnableInvest | |
|---|---|
| **Description** | Specifies the thermal generator as investment candidate [1] or omits it from the investment portfolio [0] |
| **Unit** | Boolean |
| **Parameter** | `pEnabInv(t)` |

| **InvestCost** | |
|---|---|
| **Description** | Investment cost per capacity of the thermal generator |
| **Unit** | €/MW/year |
| **Parameter** | `tThermalGen(t,'InvestCost')` |

| **FirmCapCoef** | |
|---|---|
| **Description** | Firm capacity coefficient of the thermal generator |
| **Unit** | p.u. |
| **Parameter** | `pFirmCapCoef(t)` |

| **$CO_2$Emis** | |
|---|---|
| **Description** | Specific $CO_2$ emissions of the thermal generator |
| **Unit** | $tCO_2$/MWh |
| **Parameter** | `pCO2Emis(t)` |

*7.2.3. StorageUnits*

Storage units with their properties.

| ExisUnits | |
|---|---|
| **Description** | Number of existing storage units [N] |
| **Unit** | 0 - N |
| **Parameter** | `pExisUnits(s)` |

| MaxProd | |
|---|---|
| **Description** | Maximum active power output of the storage unit |
| **Unit** | MW |
| **Parameter** | `pMaxProd(s)` |

| MinProd | |
|---|---|
| **Description** | Minimum active power output of the storage unit |
| **Unit** | MW |
| **Parameter** | `pMinProd(s)` |

| MaxCons | |
|---|---|
| **Description** | Maximum active power consumption of the storage unit (Can be 0 or empty for hydro storage power plants) |
| **Unit** | MW |
| **Parameter** | `pMaxCons(s)` |

| DisEffic | |
|---|---|
| **Description** | Discharging efficiency of the storage unit |
| **Unit** | p.u. |
| **Parameter** | `pDisEffic(s)` |

| ChEffic | |
|---|---|
| **Description** | Charging efficiency of the storage unit |
| **Unit** | p.u. |
| **Parameter** | `pChEffic(s)` |

| Qmax | |
|---|---|
| **Description** | Maximum reactive power output of the storage unit |
| **Unit** | MVAr |
| **Parameter** | `pMaxGenQ(s)` |

| Qmin | |
|---|---|
| **Description** | Minimum reactive power output of the storage unit |
| **Unit** | MVAr |
| **Parameter** | `pMinGenQ(s)` |

| InertiaConst | |
|---|---|
| **Description** | Inertia constant H of the storage unit |
| **Unit** | s |
| **Parameter** | `pInertiaConst(s)` |

| MinReserve | |
|---|---|
| **Description** | Minimum state of charge of the storage unit |
| **Unit** | p.u. |
| **Parameter** | `pMinReserve (s)` |

| IniReserve | |
|---|---|

| | |
|---|---|
| **Description** | Initial state of charge of the storage unit |
| **Unit** | p.u. |
| **Parameter** | `pIniReserve (s)` |

| **IsHydro** | |
|---|---|
| **Description** | [1] if storage unit is a hydro power plant (pumped storage power plant or storage power plant) |
| **Unit** | Boolean |
| **Parameter** | `pIsHydro(s)` |

| **OMVarCost** | |
|---|---|
| **Description** | Operation and maintenance cost of storage unit |
| **Unit** | €/MWh |
| **Parameter** | `pOMVarCost(s)` |

| **EnableInvest** | |
|---|---|
| **Description** | Specifies the storage unit as investment candidate [1] or omits it from the investment portfolio [0] |
| **Unit** | Boolean |
| **Parameter** | `pEnabInv(s)` |

| **MaxInvest** | |
|---|---|
| **Description** | Maximum number of storage units that can be invested in |
| **Unit** | 0 - N |
| **Parameter** | `pMaxinvest(s)` |

| **InvestCostPerMW** | |
|---|---|
| **Description** | Investment cost per capacity of the storage unit |

| | |
|---|---|
| **Unit** | €/MW/year |
| **Parameter** | `tStorage(s,'InvestCostPerMW' )` |

**InvestCostPerMWh**

| | |
|---|---|
| **Description** | Investment cost per energy storage capacity of the storage unit |
| **Unit** | €/MWh/year |
| **Parameter** | `tStorage(s,'InvestCostPerMWh')` |

**Ene2PowRatio**

| | |
|---|---|
| **Description** | Energy to power ratio of the storage unit. Number of hours it takes to discharge the fully charged storage unit |
| **Unit** | h |
| **Parameter** | `pE2PRatio(s)` |

**ReplaceCost**

| | |
|---|---|
| **Description** | Battery cell replacement cost |
| **Unit** | €/MWh |
| **Parameter** | `pReplaceCost(s)` |

**ShelfLife**

| | |
|---|---|
| **Description** | Shelf life of battery |
| **Unit** | Years |
| **Parameter** | `pShelfLife(s)` |

**FirmCapCoef**

| | |
|---|---|
| **Description** | Firm capacity coefficient of the storage unit |
| **Unit** | p.u. |

| | |
|---|---|
| **Parameter** | `pFirmCapCoef(s)` |

| **CDSF_alpha** | |
|---|---|
| **Description** | Cycle Depth Stress Function - alpha |
| **Unit** | - |
| **Parameter** | `pCDSF_alpha(s)` |

| **CDSF_beta** | |
|---|---|
| **Description** | Cycle Depth Stress Function - beta |
| **Unit** | - |
| **Parameter** | `pCDSF_beta(s)` |

*7.2.4. VREs*

Variable renewable energy units with their properties.

| ExisUnits | |
|---|---|
| **Description** | Number of existing VRE units [N] |
| **Unit** | 0 - N |
| **Parameter** | `pExisUnits(r)` |

| MaxProd | |
|---|---|
| **Description** | Maximum active power output of the VRE unit |
| **Unit** | MW |
| **Parameter** | `pMaxProd(r)` |

| EnableInvest | |
|---|---|
| **Description** | Specifies the VRE unit as investment candidate [1] or omits it from the investment portfolio [0] |
| **Unit** | Boolean |
| **Parameter** | `pEnabInv(r)` |

| MaxInvest | |
|---|---|
| **Description** | Maximum number of VRE units that can be invested in |
| **Unit** | 0 - N |
| **Parameter** | `pMaxinvest(r)` |

| InvestCost | |
|---|---|
| **Description** | Investment cost per capacity of VRE unit |
| **Unit** | €/MW/year |
| **Parameter** | `tRenewable(r,'InvestCost')` |

| OMVarCost | |
|---|---|
| **Description** | Operation and maintenance cost of the VRE unit |
| **Unit** | €/MWh |
| **Parameter** | `pOMVarCost(r)` |

| Qmax | |
|---|---|
| **Description** | Maximum reactive power output of the VRE unit |
| **Unit** | MVAr |
| **Parameter** | `pMaxGenQ(s)` |

| Qmin | |
|---|---|
| **Description** | Minimum reactive power output of the VRE unit |
| **Unit** | MVAr |
| **Parameter** | `pMinGenQ(s)` |

| InertiaConst | |
|---|---|
| **Description** | Inertia constant H of the VRE unit |
| **Unit** | s |
| **Parameter** | `pInertiaConst(s)` |

#### *7.2.5. VRE-profiles*

Solar and wind profiles per representative period, node, and VRE unit.

| **VRE capacity factors per representative period, hour, node, and VRE unit** | |
|---|---|
| **Description** | VRE capacity factor per rp, k, i, and VRE unit |
| **Unit** | p.u. |
| **Parameter** | `pResProfile(rp,k,i,g)` |

*7.2.6. FACTS*

FACT units and their properties.

| **ExisUnits** | |
|---|---|
| **Description** | Number of existing FACTS units [N] |
| **Unit** | 0 - N |
| **Parameter** | `pExisUnits(facts)` |

| **Qmax** | |
|---|---|
| **Description** | Maximum reactive power output of the FACTS unit |
| **Unit** | MVAr |
| **Parameter** | `pMaxGenQ(facts)` |

| **Qmin** | |
|---|---|
| **Description** | Minimum reactive power output of the FACTS unit |
| **Unit** | MVAr |
| **Parameter** | `pMinGenQ(facts)` |

| **EnableInvest** | |
|---|---|
| **Description** | Specifies the FACTS unit as investment candidate [1] or omits it from the investment portfolio [0] |
| **Unit** | Boolean |
| **Parameter** | `pEnabInv(facts)` |

| **MaxInvest** | |
|---|---|
| **Description** | Maximum number of FACTS units that can be invested in |
| **Unit** | 0 - N |
| **Parameter** | `pMaxinvest(facts)` |

| **InvestCost** | |
|---|---|
| **Description** | Investment cost per capacity of FACTS unit |
| **Unit** | €/MVar/year |
| **Parameter** | `tFACTS(facts,'InvestCost')` |

#### *7.2.7. Inflows*

Inflow for hydro generator per representative period, hour, and node.

| **Inflows per representative period, hour, and node.** | |
|---|---|
| **Description** | Inflows for hydro units per representative period, hour, and node |
| **Unit** | MWh |
| **Parameter** | `pInflows(rp,k,g)` |

*7.2.8. Weights*

Number of occurrence (weight) of representative period and hourly weight for representative period.

| Representatives periods weight | |
|---|---|
| **Description** | Number of occurrence (weight) of representative period |
| **Unit** | N |
| **Parameter** | `pWeight_rp(rp)` |
| **Hourly weight for representative period** | |
| **Description** | Hourly weight for representative period |
| **Unit** | N |
| **Parameter** | `pWeight_k (k)` |

*7.2.9. Hindex*

Relation among chronological periods, representative periods, and hours within representative period.

| Relation among time indices | |
|---|---|
| **Description** | Relation among periods, representative periods, and hours within representative period. Indicates which chronological periods $p$ are represented by representative period $rp$ and hours $k$ within $rp$. Usually the result of a clustering procedure. |
| **Unit** | - |
| **Source** | Output of clustering procedure |
| **Parameter** | `hindex(p,rp,k)` |

#### *7.2.10. DSM-DelayTime*

Delay time for DSM shifting for different sectors. Delay time for load shifting per sector (backward or forward).

| **Delay time** | |
|---|---|
| **Description** | Hours of load shifting per sector and representative period |
| **Unit** | h |
| **Parameter** | `pDSMDelayTime(sec,rp)` |

*7.2.11. DSM-Max*

Represents the DSM shifting profile of total demand that can be used for upward or downward DSM shifting.

| **Upward shifting** | |
|---|---|
| **Description** | Maximum upward shifting DSM per representative period, hour, node, and sector |
| **Unit** | p.u. |
| **Parameter** | `pMaxUpDSM(rp,k,i,sec)` |
| **Downward shifting** | |
| **Description** | Maximum downward shifting DSM per representative period, hour, node, and sector |
| **Unit** | p.u. |
| **Parameter** | `pMaxDnDSM(rp,k,i,sec)` |

*7.2.12. DSM-shift*

Demand-side management shifting cost.

| DSM shifting cost | |
|---|---|
| **Description** | Cost profile for DSM load shifting per representative period and node |
| **Unit** | €/MWh |
| **Parameter** | `pDSMShiftCost(rp,k,i)` |

*7.2.13. DSM-shed*

Demand-side management shedding segments.

| | |
|---|---|
| **ShedPercentage** | |
| **Description** | Percentage of load applicable for shedding. |
| **Unit** | p.u. |
| **Parameter** | `pDSMShedRatio(seg)` |
| **ShedPenalty** | |
| **Description** | Cost of price-responsive demand side management |
| **Unit** | €/MWh |
| **Parameter** | `pDSMShedCost(seg)` |

*7.2.14. Network*

Information about the network topology and the properties of power lines or transformers. Transformers can be specified in the same way as power lines (including tap angle and tap ratio). For the sake of simplicity, we will just refer to power lines here.

| **From bus i** | |
|---|---|
| **Description** | Starting node of a power line |
| **Unit** | - |
| **Set** | i |
| **To bus j** | |
| **Description** | End node of a power line |
| **Unit** | - |
| **Set** | j |
| **Circuit ID** | |
| **Description** | Indicates the circuit of a power line connecting two nodes (to differentiate parallel power lines) |
| **Unit** | c1-cN |
| **Set** | c |
| **InService** | |
| **Description** | Indicates if a power line is in service [1], or not [0]. Candidate power lines have to be in service [1] in order to be considered as such. |
| **Unit** | Boolean |
| **Parameter** | `tNetwork(i,j,c,'InService')` |
| **R** | |
| **Description** | Resistance R of a power line |
| **Unit** | p.u. |

| | |
|---|---|
| **Parameter** | `pRline` |
| | **X** |
| **Description** | Reactance X of a power line |
| **Unit** | p.u. |
| **Parameter** | `pXline` |
| | **Bc** |
| **Description** | Branch charging susceptance Bc of a power line |
| **Unit** | p.u. |
| **Parameter** | `pBcline` |
| | **TapAngle** |
| **Description** | Shift angle of a transformer |
| **Unit** | ° |
| **Parameter** | `pAngle` |
| | **TapRatio** |
| **Description** | Tap ratio of a transformer |
| **Unit** | p.u. |
| **Parameter** | `pRatio` |
| | **Pmax** |
| **Description** | Maximum active power capacity of a power line |
| **Unit** | MW |
| **Parameter** | `pPmax` |
| | **FixedCost** |
| **Description** | Investment cost for a candidate power line |
| **Unit** | M€ |
| **Parameter** | `tNetwork(i,j,c,'FixedCost')` |
| | **FxCharageRate** |

| | |
|---|---|
| **Description** | Annualization factor of the investment cost for a candidate power line |
| **Unit** | p.u. |
| **Parameter** | `tNetwork(i,j,c,'FxChargeRate')` |

*7.2.15. BusInfo*

Nodes and their properties.

| **BaseVolt** | |
|---|---|
| **Description** | Base voltage at bus i |
| **Unit** | kV |
| **Parameter** | `pBusBaseV(i)` |
| **maxVolt** | |
| **Description** | Maximum voltage at bus i (factor that gets multiplied by BaseVolt) |
| **Unit** | p.u. |
| **Parameter** | `pBusMaxV(i)` |
| **minVolt** | |
| **Description** | Minimum voltage at bus i (factor that gets multiplied by BaseVolt) |
| **Unit** | p.u. |
| **Parameter** | `pBusMinV(i)` |
| **Bs** | |
| **Description** | Suceptance B connected at bus i |
| **Unit** | p.u. |
| **Parameter** | `pBusB(i)` |
| **Gs** | |
| **Description** | Conductance G connected at bus i |
| **Unit** | p.u. |
| **Parameter** | `pBusG(i)` |
| **powerFactor** | |
| **Description** | Power factor at bus i |
| **Unit** | p.u. |

**Parameter** `pBus_pf(i)`

---

*7.3. H2-Sector*

The sheets that contain input data for the hydrogen sector are explained in this section.

*7.3.1. H2_Indices*

Definition of indices for the hydrogen sector.

| **h2sec** | |
|---|---|
| **Description** | Sectors of hydrogen demand |
| **Set** | `h2sec` |

| **h2tec** | |
|---|---|
| **Description** | Hydrogen technologies |
| **Set** | `h2tec` |

| **h2u** | |
|---|---|
| **Description** | Set of all hydrogen units |
| **Set** | `h2u` |

| **h2i** | |
|---|---|
| **Description** | Nodes for hydrogen pipeline network (h2j is the alias of h2i) |
| **Set** | `h2i` |

| **h2gh2i** | |
|---|---|
| **Description** | Unit h2g connected to hydrogen node h2i |
| **Set** | `h2gh2i(h2u,h2i)` |

| **h2gi(h2u,i)** | |
|---|---|
| **Description** | Unit h2u connected to electrical node i |
| **Set** | `h2gi(h2u,i)` |

| **h2uh2tec(h2u,h2tec)** | |
|---|---|
| **Description** | Relation among h2 units and technologies |

**Set** `h2uh2tec(h2u,h2tec)`

*7.3.2. H2_Demand*

Hourly hydrogen demand for each hydrogen node, representative period, and hydrogen sector.

| H2_Demand | |
|---|---|
| **Description** | Hourly hydrogen demand for each hydrogen node, representative period, and hydrogen sector |
| **Unit** | kg |
| **Parameter** | `pH2Demand(rp,k,h2i,h2sec)` |

*7.3.3. H2_GenUnits*

Hydrogen production units with their properties.

| ExisUnits | |
|---|---|
| **Description** | Number of existing hydrogen production units [N] |
| **Unit** | Boolean |
| **Parameter** | `pH2ExisUnits(h2u)` |

| MaxCons | |
|---|---|
| **Description** | Maximum active power consumption of hydrogen production unit |
| **Unit** | MW |
| **Parameter** | `pH2MaxCons(h2u)` |

| OMVarCost | |
|---|---|
| **Description** | Operation and maintenance variable cost of hydrogen production unit. |
| **Unit** | % of CAPEX/year |
| **Parameter** | `pH2OMPercent(h2u)` |

| PowerFactor | |
|---|---|
| **Description** | Power factor of hydrogen production unit |
| **Unit** | p.u. |
| **Parameter** | `pH2_pf(h2u)` |

| EnableInvest | |
|---|---|
| **Description** | Specifies the hydrogen production unit as investment candidate [1] or omits it from the investment portfolio [0] |
| **Unit** | Boolean |
| **Parameter** | `tH2GenUnits(h2u,'EnableInvest')` |

| InvestCost | |
|---|---|
| **Description** | Investment cost of hydrogen production unit |

| | |
|---|---|
| **Unit** | €/MW/year |
| **Parameter** | `pH2InvestCost(h2g)` |

**MaxInvest**

| | |
|---|---|
| **Description** | Maximum number of hydrogen production units that can be invested in |
| **Unit** | 0 - N |
| **Parameter** | `pH2MaxInvest(h2g)` |

#### *7.3.4. H2_Network*

Information about the pipelines and their properties.

| **From bus h2i** | |
|---|---|
| **Description** | Starting node of a hydrogen pipeline |
| **Unit** | - |
| **Set** | h2i |
| **To bus h2j** | |
| **Description** | End node of a hydrogen pipeline |
| **Unit** | - |
| **Set** | h2j |
| **InService** | |
| **Description** | Indicates if a hydrogen pipeline is in service [1], or not [0]. Candidate hydrogen pipelines have to be in service [1] in order to be considered as such. |
| **Unit** | Boolean |
| **Parameter** | `tH2Network(h2i,h2j,'InService')` |
| **Fmax** | |
| **Description** | Maximum hydrogen flow through hydrogen pipeline |
| **Unit** | kg |
| **Parameter** | `pH2Fmax` |

## 8. Outputs

LEGO is a cost minimization optimization model, which - depending on the modules of interest that were chosen - yields a wide variety of outputs (via its primal variables). An exhaustive list of the model variables is provided in the appendix, and a list of all outputs that are automatically generated is provided in the remainder of this section. However, LEGO can also employ dual variables for price information and it can hence be used for economic assessments of particular units or technologies, which makes it a useful tool to study impacts of climate policies on market participants.

Just some quick words about duality and price information. How do we use LEGO to obtain price information and to calculate generator profits? If integrality in LEGO is relaxed, then the model becomes a linear program (LP) and dual variables are uniquely defined due to strong duality of LPs. However, in a MIP framework this is not the case. A common practice to obtain prices for MIP models, which has also been adopted in this paper, is to: run the MIP; then fix all integer variables to their optimal value; and re-run the model as an LP. This method determines equilibrium prices for a given solution. With this in mind, prices are obtained as the dual variables of the corresponding constraints, i.e., the spot market price is the dual variable of the demand balance constraint, reserve market prices are obtained as dual variables of reserve constraints etc.

In that sense, we can calculate profits of each generator and differentiate among different cost and revenue streams as follows:

- Spot market revenues minus spot market purchases (the latter can only occur for storage units). These revenues are calculated using the spot market price, i.e., the dual variable of the demand balance constraint.

- Reserve market revenues minus reserve market O&M cost. Reserve prices are obtained as dual variables of reserve constraints.

- Renewable quota payments. LEGO has a constraint that forces a certain amount of renewable penetration (to mimic policy goals). Its dual can be used to determine corresponding payments.

- Firm capacity payments (should there be any). Firm capacity price is obtained as dual of firm capacity constraint.
- Minus operating costs (variable and fixed).
- Minus investment costs.

### *8.1. General*

The following subsections provide a list with all the names of the sheets that contain outputs; and for each sheet the output data with a short description and unit is described. Note that the title of each subsection coincides with the name of the corresponding data sheet in the input Excel.

*8.1.1. Summary*

The summary sheet provides a brief overview of model results, including model statistics and information on the power system, hydrogen system, $CO_2$ emissions, and policies.

| | |
|---|---|
| **Model Statistics** | |
| **Description** | Overview of model statistics including: objective function value; total CAPEX; total OPEX; CPU time (model generation and solution time); number of variables; number of discrete variables; number of equations; number of nonzero elements; best possible solution for MIP; results for regret calculation; SOCP or DC; RoCoF or MinInert; SOCP mean error |
| **Power System** | |
| **Description** | Summary of the main results of the power system: Total system demand; total renewable + storage production; total renewable curtailment; total thermal production; actual green production; actual thermal production; actual $CO_2$ emissions; thermal investment; renewable investment; storage investment; transmission line investment; energy non-supplied |
| **Hydrogen System** | |
| **Description** | Summary of the main results of the hydrogen system: H2 non-supplied |
| **CO2 Emissions** | |
| **Description** | Summary of the main results of $CO_2$ emissions: Budget $CO_2$ emissions; $CO_2$-target overshoot |
| **Policies** | |

| | |
|---|---|
| **Description** | Summary of the main results of policies: Cost renewable quota; payment firm capacity |

### *8.1.2. TotalEn*

| **TotalEn** | |
|---|---|
| **Description** | Total produced energy per technology and node as well as consumption, demand, and energy non-supplied per node. |
| **Unit** | GWh, GVarh |
| **Parameter** | `pTecProd` |
| **Variables** | `vGenP`, `vGenQ`, `vConsump`, `pDemandP`, `pDemandQ`, `vLineP`, `vLineQ`, `vDSM_Shed`, `vPNS` |

### *8.1.3. GenInvest*

| **GenInvest** | |
|---|---|
| **Description** | Investment in candidate power plants. (If run as rMIP this can be fractions of the installed capacity of thermal power plants as well.) |
| **Unit** | MW, MVar |
| **Parameter** | `pGenInvest` |
| **Variables** | `vGenInvest` |

### *8.1.4. H2Invest*

| **H2Invest** | |
|---|---|
| **Description** | Investment in hydrogen production units. |
| **Unit** | MW |
| **Parameter** | `pH2Invest` |
| **Variables** | `vH2Invest` |

*8.1.5. CDSF*

| **CDSF** | |
|---|---|
| **Description** | Objective function for cycle aging costs of storage units [M€]; annual life loss from cycling [%]; annual prorated cycle aging cost [M€]; and battery life expectancy [year] |
| **Unit** | M€, %, M€, year |
| **Parameter** | `pResulCDSF` |

*8.1.6. UC*

| **UC** | |
|---|---|
| **Description** | Unit commitment of thermal power plants per hour |
| **Unit** | When run as MIP: Boolean; When run as rMIP: continuous between 0 - 1 |
| **Parameter** | `pCommit` |
| **Variables** | `vCommit` |

*8.1.7. GenP*

| **GenP** | |
|---|---|
| **Description** | Active power generation of generation units per hour |
| **Unit** | MW |
| **Parameter** | `pGenP` |
| **Variables** | `vGenP` |

*8.1.8. GenQ*

| GenQ | |
|---|---|
| **Description** | Reactive power generation of generation units per hour |
| **Unit** | MVar |
| **Parameter** | `pGenQ` |
| **Variables** | `vGenQ` |

*8.1.9. Charge*

| Charge | |
|---|---|
| **Description** | Charging power of the storage unit per hour |
| **Unit** | MW |
| **Parameter** | `pChrP` |
| **Variables** | `vConsump` |

*8.1.10. Curtail*

| Curtail | |
|---|---|
| **Description** | Curtailment of VRE units per hour (k) and per representative period (rp) |
| **Unit** | MW |
| **Parameter** | `pCurtP_k`, `pCurtP_rp` |
| **Variables** | `vGenInvest`, `vGenP` |

*8.1.11. StIntra*

| **StIntra** | |
|---|---|
| **Description** | Intra-period state of charge of storage units |
| **Unit** | % |
| **Parameter** | `pStIntra` |
| **Variables** | `vStIntraRes` |

*8.1.12. StLevel*

| **StLevel** | |
|---|---|
| **Description** | Inter-period state of charge of storage units |
| **Unit** | % |
| **Parameter** | `pStLevel` |
| **Variables** | `vStInterRes` |

*8.1.13. MC*

| **MC** | |
|---|---|
| **Description** | Electricity prices per node and hour |
| **Unit** | €/MWh |
| **Parameter** | `pSRMC` |

*8.1.14. InertDual*

| **InertDual** | |
|---|---|
| **Description** | Dual variable of inertia constraint per representative period (rp) and hour (k) inside rp |
| **Unit** | €/s |
| **Parameter** | `pInertDual` |

*8.1.15. RoCoF*

| **RoCoF** | |
|---|---|
| **Description** | Actual system inertia per representative period (rp) and hour (k) inside rp |
| **Unit** | s |
| **Parameter** | `pActualSysInertia` |

*8.1.16. LineP*

| **LineP** | |
|---|---|
| **Description** | Active power flow per power line, circuit, representative period (rp), and hour (k) inside rp |
| **Unit** | MW |
| **Parameter** | `pLineP` |
| **Variables** | `vLineP` |

*8.1.17. LineQ*

| **LineQ** | |
|---|---|
| **Description** | Reactive power flow per powe line, circuit, representative period (rp), and hour (k) inside rp |
| **Unit** | MVar |
| **Parameter** | `pLineQ` |
| **Variables** | `vLineQ` |

*8.1.18. Volt*

| Volt | |
|---|---|
| **Description** | Voltage of bus/node per representative period (rp) and hour (k) inside rp |
| **Unit** | p.u. |
| **Parameter** | `pVoltage` |

*8.1.19. Angle*

| Angle | |
|---|---|
| **Description** | Voltage angle of bus/node per representative period (rp) and hour (k) inside rp |
| **Unit** | ° |
| **Parameter** | `pTheta` |
| **Variables** | `vTheta` |

*8.1.20. Profits*

| Profits | |
|---|---|
| **Description** | Spot market revenues; spot market costs; reserve market revenues; reserve market costs; O&M costs; investment costs; RES quota payment/cost; firm capacity payments; total profits |
| **Unit** | M€ |
| **Parameter** | `pEconomicResults` |

*8.1.21. DSM*

| DSM | |
|---|---|

| | |
|---|---|
| **Description** | Results of DSM per segment, node, representative period (rp), and hour (k) inside rp |
| **Unit** | GW |
| **Parameter** | `pResultDSM` |
| **Variables** | `vDSM_Up`, `vDSM_Dn`, `vDSM_Shed` |

*8.1.22. H2price*

| **H2price** | |
|---|---|
| **Description** | Price of hydrogen |
| **Unit** | €/kg |
| **Parameter** | `pH2price` |

#### *8.1.23. H2Prod*

| H2Prod | |
|---|---|
| **Description** | Hydrogen production per unit, representative period (rp), and hour (k) inside rp |
| **Unit** | kg |
| **Parameter** | `pH2Prod` |
| **Variable** | `vH2Prod` |

#### *8.1.24. H2Cons*

| H2Cons | |
|---|---|
| **Description** | Power consumption per hydrogen unit, representative period (rp), and hour (k) inside rp |
| **Unit** | MW |
| **Parameter** | `pH2Cons` |
| **Parameter** | `vH2Consump` |

#### *8.1.25. H2ns*

| H2ns | |
|---|---|
| **Description** | Hydrogen non-supplied per hydrogen node, sector, representative period (rp), and hour (k) inside rp |
| **Unit** | kg |
| **Parameter** | `pH2ns` |
| **Parameter** | `vH2NS` |

## 9. How to construct new cases

The easiest way to construct a new case is to take one of the existing data files and alter it, simply because in the Excel files, there are certain ranges already defined that are directly imported into GAMS. Below you will find a couple of changes you can do WITHOUT having to have prior knowledge in GAMS.

- You can change the data at hand at your will. For example, updating fuel costs, investment costs, efficiency factors, etc. Just be careful when doing that because some cells contain formulas, and you don't want to change those (or maybe you do).

- Adding a new element to an existing index/set. In the Index sheet, simply insert a row where convenient (within the existing range). The name of the element cannot coincide with an existing element though. Depending on what you add, the corresponding data sheets need to be updated. Example 1: how to add a new thermal generator? You add a row belonging to index $g$ where it says thermal units; then add a row for $gi(g,i)$ and assign the new generator to a bus; under $gtec(g,tec)$ assign a technology to the generator; then add a corresponding row in data sheet ThermalGen where you put the name of your new generator and fill out the required data. Example 2: adding a renewable generator. In the index sheet it is the same procedure; but then you have add a row with corresponding data in sheet VRES; in sheet VRES-profile you will have to include several new columns (depending on temporal granularity) to contain the hourly profile of the renewable generator you added. Example 3: how to add a new power line. If the buses already exist, simply go to sheet Network, and add a new line within the existing range there; then specify the data of this new line (start bus, end bus, reactance, resistance etc.); if you want the line to be a candidate line then you also have to specify a FixedCost and a FxChargeRate (which represents investment costs).

- Erasing an existing element from an index. The procedure is similar to adding an element. One has to go through data sheets erasing rows and corresponding columns. Example: you want to erase a storage unit. In the Indices sheet, erase the row containing the storage unit in question; erase the corresponding rows of *gtec* and *gi* that contain your unit; in sheet StorageUnits erase the entire row that contains the unit you want to erase.

- Changing the temporal representation. Here you have to be a little more careful. If you want to model something with chronological hours, then take the "hourly" data case as an example and alter it by simply adding more elements to index $p$ (or erasing them if you want fewer units); you will also have to alter all data that uses the temporal representation, i.e. demand, renewable profiles, hydro inflows, weights, Hindex and H2 demand. We explain how to change such a table by the example of electricity demand: go to sheet Demand; if you look for Excel range "demand" you will see the range that will be exported directly to GAMS; then, add rows (within the existing range) because you want to increase $p$, and then fill out the corresponding data. If you want to use representative days, then either you know which days that are, or you have run some clustering algorithm at some point (which would give you centroids, the list of which hours belong to which representative period, etc.). Let us say that you want to create a 2 representative day sheet, then you take the existing 7 representative day sheet, reduce $rp$ in the Indices sheet to 2, and in the Demand, VRES-profiles, etc. sheets you start erasing all columns/rows that mention $rp3 - rp7$. Doing so will give you at least the correct data sheet structure. Now you just have to replace the old data with the centroids (from your clustering algorithm) for demand and renewable profiles etc.

There are many more things you could do to change LEGO, but for those you will have to know how to code in GAMS.

- Adding a new module: if you want to add a new module that does not already exist, for example, the heating sector, then you should create new sheets in Excel that define indices and corresponding parameters (just make sure your names are unique); then define ranges in Excel in order to be able to export them to GAMS; if you want to know how data input is handled in the LEGO GAMS file just pick a particular range and follow its path in the GAMS file. In general, all ranges are stored in tables whose name starts with "t". Then, they are imported into model parameters, all of which start with "p". Then, you can create new constraints using your new indices and parameters. You just need to be careful when you establish the links between new and old modules. For example, when adding the heat sector, where would that influence the power sector? For example: is there a share of heating demand that is provided by heat pumps, which increases electric power demand and affects the existing power balance?

## 10. APPENDIX A: Nomenclature

Acronyms:

| | |
|---|---|
| AC/DC | Alternating/direct current |
| BESS | Battery energy storage system |
| CCGT/OCGT | Combined cycle/Open cycle gas turbine |
| CTS | Commerce, trade and services |
| DSM | Demand side management |
| ESM | Energy system model |
| EV | Electric vehicle |
| GEP/TEP | Generation/Transmission expansion planning |
| LEGO | Low-carbon expansion generation optimization |
| LP | Linear program |
| MIP | Mixed integer program |
| rMIP | Relaxed mixed integer program |
| MIQCP | Mixed integer quadratically constrained program |
| OPF | Optimal power flow |
| O&M | Operations and maintenance |
| RoCoF | Rate of change of frequency |
| SOCP | Second-order cone program |
| UC | Unit commitment |

Below we include the nomenclature used in the mathematical formulation of the LEGO model, which is also presented in the appendix. Please note that, this mathematical formulation (and the corresponding nomenclature) does not exactly coincide with the nomenclature in the GAMS code for reasons of simplicity and legibility.

Indices:

| | |
|---|---|
| $p$ | Time periods (usually hours) |
| $rp$ | Representative periods (usually days) |

| | |
|---|---|
| $k$ | Time periods within a representative period |
| $\Gamma(p, rp, k)$ | Mapping of periods with representative periods $rp$ and $k$ |
| $g$ | Generating units |
| $t(g)$ | Subset of thermal generation units |
| $s(g)$ | Subset of storage generation units |
| $r(g)$ | Subset of renewable generation units |
| $v(g)$ | Subset of units that provide virtual inertia |
| $sec$ | Sectors for demand-side management shifting |
| $seg$ | Segments for price-responsive DSM |
| $facts(g)$ | Subset of FACTS as reactive power source |
| $i, j, ii$ | Bus of transmission network |
| $iws$ | Transmission busses without slack bus |
| $c$ | Circuit in transmission network |
| $ijc(i, j, c)$ | Transmission line connecting nodes $i$,$j$ with $c$ |
| $ijce(i, j, c)$ | Existing transmission line connecting nodes $i$,$j$ with $c$ |
| $ijcc(i, j, c)$ | Candidate transmission line connecting nodes $i$,$j$ with $c$ |
| $line(i, j)$ | Indicates if a line exists between nodes $i$ and $j$ |
| $gi(g, i)$ | Generator $g$ connected to node $i$ |
| $h2sec$ | Sectors of hydrogen demand |
| $h2g$ | Subset of hydrogen generating units |
| $h2i, h2j$ | Node of hydrogen network |
| $h2gh2i(h2g, h2i)$ | Unit $h2g$ connected to node $h2i$ |
| $h2gi(h2g, i)$ | Unit $h2g$ connected to bus $i$ |

Parameters:

| | |
|---|---|
| $D^P_{rp,k,i}$ | Active power demand (GW) |
| $D^Q_{rp,k,i}$ | Reactive power demand (GW) |
| $\eta^{DIS}_g$ | Discharge efficiency of unit (p.u.) |
| $\eta^{CH}_g$ | Charge efficiency of unit (p.u.) |
| $B_i$ | Susceptance connected at bus $i$ (p.u.) |

| | |
|---|---|
| $B_{i,j,c}$ | Line susceptance (p.u.) |
| $Bc_{i,j,c}$ | Branch charging susceptance (p.u.) |
| $G_i$ | Conductance connected at bus $i$ (p.u.) |
| $G_{i,j,c}$ | Line conductance (p.u.) |
| $SB$ | Base power (MVA) |
| $R_i$ | Tan(arccos(pf)) = Q/P at bus $i$ (p.u.) |
| $W_{rp}^{RP}$ | Weight of the representative period (h) |
| $W_k^K$ | Weight of each $k$ within the representative period (h) |
| $C^{ENS}$ | Cost of energy non-served (M€/GWh) |
| $C^{DMS,S}$ | Cost of DSM shedding (M€/GWh) |
| $C^{DMS,-}$ | Cost of DSM shifing (M€/GWh) |
| $C_g^{SU}$ | Start-up cost of unit (M€) |
| $C_g^{UP}$ | Commitment cost of unit (M€/h) |
| $C_g^{VAR}$ | Variable cost of energy (M€/GWh) |
| $C_g^{OM}$ | Operation and maintenance cost (M€/GWh) |
| $C_g^{INV}$ | Investment cost (M€/GW/y) |
| $C_{i,j,c}^{L,INV}$ | Line investment cost (M€/GW/y) |
| $C^{RES+}$ | Reserve-up cost (p.u.) |
| $C^{RES-}$ | Reserve-down cost (p.u.) |
| $RES^+$ | System reserve-up requirement (p.u.) |
| $RES^-$ | System reserve-down requirement (p.u.) |
| $\underline{P}_g$ | Technical minimum of unit (GW) |
| $\overline{P}_g$ | Technical maximum of unit (GW) |
| $EU_g$ | Indicator of existing unit (integer) |
| $RU_g$ | Ramp-up limit of unit (GW) |
| $RD_g$ | Ramp-down limit of unit (GW) |
| $MOW$ | Moving window for long-term storage (h) |
| $PF_{rp,k,i,r}$ | Renewable profile per unit and node (p.u.) |
| $\underline{R}_s$ | Minimum reserve of storage unit (p.u.) |
| $M_{rp,k,s}^{ch/d}$ | Upper bound on charge and discharge (GW) |

| | |
|---|---|
| $InRes_{s,p}$ | Initial reserve (GWh) |
| $IF_{rp,k,s}$ | Inflows (GWh) |
| $\kappa$ | Minimum clean (s+r) production (p.u.) |
| $ISF_{i,j,c,ii}$ | Injection Shift Factors (p.u.) |
| $\overline{T}_{i,j,c}$ | Transmission line limit (GW) |
| $\overline{A}_{i,j,c}$ | Apparent power transfer limit (MVA) |
| $\Delta$ | Maximum angle difference (rad) |
| $\overline{X}_g$ | Maximum amount of units to be built (integer) |
| $\overline{X}^{H2}_{h2g}$ | Maximum amount of hydrogen units to be built (integer) |
| $\overline{X}^{L}_{i,j,c}$ | Maximum amount of transmission lines to be built $\in \{0, 1\}$ |
| $D^{H2}_{rp,k,h2i,h2sec}$ | Hydrogen demand per sector (t) |
| $HPE_{h2g}$ | Hydrogen per unit of energy (t/GWh) |
| $C^{H2,INV}_{h2g}$ | Investment cost for hydrogen unit (M€/GW/y) |
| $\overline{P}^{E}_{h2g}$ | Technical maximum of hydrogen unit (GW) |
| $EU^{H2}_{h2g}$ | Indicator of existing hydrogen units (integer) |

Variables:

| | |
|---|---|
| $p_{rp,k,g}$ | Real power generation of the unit (GW) |
| $\hat{p}_{rp,k,g}$ | Real power generation above the technical minimum (GW) |
| $q_{rp,k,g}$ | Reactive power generation of the unit (Gvar) |
| $cs_{rp,k,g}$ | Consumption of the unit (GW) |
| $pns_{rp,k,i}$ | Power non-served (GW) |
| $f^{P}_{rp,k,i,j,c}$ | Real power flow of line $ijc$ (GW) |
| $f^{Q}_{rp,k,i,j,c}$ | Reactive power flow of line $ijc$ (Gvar) |
| $so^{cii}_{rp,k,i}$ | Auxiliary $cii$ variable for SOCP formulation (p.u.) |
| $y_{rp,k,g}$ | Startup decision of the unit (integer) |
| $z_{rp,k,g}$ | Shutdown decision of the unit (integer) |
| $u_{rp,k,g}$ | Dispatch commitment of the unit (integer) |
| $x_g$ | Investment in generation capacity (integer) |
| $b^{ch/d}_{rp,k,s}$ | Indicator if storage is charging or discharging (binary) |

| | |
|---|---|
| $sp_{rp,k,s}$ | Spillages or curtailment (GWh) |
| $res^{+}_{rp,k,g}$ | Secondary reserve up allocation (GW) |
| $res^{-}_{rp,k,g}$ | Secondary reserve down allocation (GW) |
| $inter_{p,s}$ | Inter-period storage reserve or state of charge (GWh) |
| $p^{H2}_{rp,k,h2g}$ | Hydrogen generation of the unit (t) |
| $cs^{E}_{rp,k,h2g}$ | Electric power consumption of the unit (GW) |
| $h2ns_{rp,k,h2i}$ | Hydrogen non-served (t) |
| $x^{H2}_{h2g}$ | Investment in hydrogen capacity (integer) |

## 11. APPENDIX B: Mathematical Model Formulation

In this section we present the mathematical formulation of the LEGO model. However, since LEGO was based on a preliminary model version [8]. The novelties of the model contain: the fact that LEGO does now include demand side management (DSM) discussed in section 11.2; transmission expansion planning in a DC- and an AC-OPF setting described in section 11.3; a simple representation of the hydrogen sector via an electrolyzer presented in section 11.4. The other constraints are presented subsequently.

For the sake of clarity, index $g$ represents all generating units as a whole (both existing and candidate units) in the power sector, and sub-indices $t, r, s$ are thermal, renewable and storage units. On the other hand, index $h2g$ describes hydrogen generating units. Please note that, the nomenclature of the mathematical formulation does not coincide with the nomenclature of the GAMS code, simply because the code follows some conventions and coding standards, which would lead to mathematical formulations that are illegible in a paper format.

### 11.1. Objective function and standard constraints

In this section we present the objective function and discuss some other standard constraints. The objective function (1a) represents total system cost as: thermal production cost (start-up cost, commitment cost, and variable cost); potential cost for non-supplied energy; battery degradation cost[9]; cost of providing upward and downward secondary reserve by thermal and storage units; cost of providing demand side management (shifting or shedding demand); generation expansion investment costs for building new units, transmission expansion costs for building new lines, and investment (and O&M) costs of building new hydrogen units; finally, we also consider a potential penalty for hydrogen non-served. Constraint (1b) represents the upper and lower bounds of non-supplied energy; and, (1c) defines investment variables as non-negative integers and establishes an upper bound introduced by parameters $\overline{X}$.

$$\begin{aligned}
min \sum_{rp,k} W^{RP}_{rp} W^{K}_{k} \Big( \sum_{t} (C^{SU}_{t} y_{rp,k,t} + C^{UP}_{t} u_{rp,k,t} + C^{VAR}_{t} p_{rp,k,t}) \\
+ \sum_{r} C^{OM}_{r} p_{rp,k,r} + \sum_{s} C^{OM}_{s} p_{rp,k,s} + \sum_{i} C^{ENS} pns_{rp,k,i} \Big) \\
+ \sum_{rp,k,cdsf(s),a} W^{RP}_{rp} W^{K}_{k} C^{CDSF}_{s,a} dis_{rp,k,s,a} \\
+ \sum_{rp,k} W^{RP}_{rp} W^{K}_{k} \Big( \sum_{t} (C^{VAR}_{t} C^{RES+} res^{+}_{rp,k,t} + C^{VAR}_{t} C^{RES-} res^{-}_{rp,k,t}) \\
+ \sum_{s} (C^{OM}_{s} C^{RES+} res^{+}_{rp,k,s} + C^{OM}_{s} C^{RES-} res^{-}_{rp,k,s}) \Big) \\
+ \sum_{rp,k} W^{RP}_{rp} W^{K}_{k} \Big( \sum_{i,sec} C^{DSM,-} dsm^{-}_{rp,k,i,sec} + \sum_{i,seg} C^{DSM,S} dsm^{S}_{rp,k,i,seg} \Big) \\
+ \sum_{g} C^{INV}_{g} x_{g} + \sum_{ijcc(i,j,c)} C^{L,Inv}_{i,j,c} x^{L}_{i,j,c} + \sum_{h2g} (C^{H2,INV}_{h2g} + C^{H2,OM}_{h2g}) x^{H2}_{h2g} \\
+ \sum_{rp,k,hi} W^{RP}_{rp} W^{K}_{k} C^{HNS} h2ns_{rp,k,hi} \quad (1a)
\end{aligned}$$

$$0 \leq pns_{rp,k,i} \leq D^{P}_{rp,k,i} \quad \forall rp,k,i \quad (1b)$$

$$x_{g} \in \mathbb{Z}^{+,0}, x_{g} \leq \overline{X}_{g} \quad \forall g \quad (1c)$$

Other standard constraints, that have been previously presented in [8] and therefore do not constitute an original contribution of this paper, are briefly

[9]If no battery degradation is considered then this term is simply zero.

mentioned here and listed in the appendix for completeness. For a detailed explanation of these constraints, the reader is referred to [8]. In particular, the constraints regarding thermal power plant operations are given by (7), and the constraints that are specific for storage technologies are given by (8).

Constraints (2) are policy constraints and related to variable renewable energy sources. In particular, (2a) corresponds to lower and upper bounds on renewable production. In order to be able to control renewable penetration, a system-wide constraint that limits thermal production to at most $(1-\kappa)$ percent of total system demand (2b) is also imposed[10], thereby implicitly forcing $\kappa$ percent clean production. Inspired by [9], we have introduced a firm capacity constraint (2c) where $FCC_g$ is the firm capacity coefficient[11] by technology and $FCP$ is the percentage of firm capacity required by the system (110% here) both have been taken from [9]. In the rMIP framework, the dual of this constraint yields a firm capacity price in €/MW of firm capacity, which is used in firm capacity payments.

$$0 \leq p_{rp,k,r} \leq \overline{P}_r PF_{rp,k,r}(x_r + EU_r) \quad \forall rp,k,r \tag{2a}$$

$$\sum_{rp,k,t} W_k^K W_{rp}^{RP} p_{rp,k,t} \leq (1-\kappa) \sum_{rp,k,i} W_k^K W_{rp}^{RP} D_{rp,k,i}^P \tag{2b}$$

$$\sum_{g} FCC_g \overline{P}_g (x_g + EU_g) \geq FCP\ D^{peak} \tag{2c}$$

*11.2. Demand-side management*

One novelty in LEGO is the fact that it can now account for demand-side management (DSM). In particular, we consider two different types of DSM: demand shifting within a given range of hours and demand shedding. In order

[10]Note that if parameter $\kappa$ is set to zero, this constraint is relaxed

[11]This factor describes the percentage of the installed capacity for each technology that is considered firm. In this context, firm means the capacity available for production or transmission which can be (and in many cases must be) guaranteed to be available at a given time. These coefficients are almost 100% for dispatchable technologies, and usually much lower for intermittent VREs where they could also be a function of the system portfolio.

to do so, we introduce two new indices, i.e., *sec*, which represents the different sectors where DSM shifting is considered (such as electric vehicles, residential heating or residential washing/drying load), and *seg*, which corresponds to the different segments of price-responsive DSM. The DSM shifting that we represent considers that a particular load can be deferred within a particular amount of time. For example, the load of washing and drying at 10am can be deferred for at most 6 hours, so until 4pm but not more. In order to establish the correct relation among inter-day hours where DSM shifting is allowed, we define the dynamic set $dkk_{rp,k,k,sec}$.

In (3) we present the corresponding DSM constraints. Note that DSM also has an impact on the demand balance equations, which are discussed in detail in section 11.3. In order to represent the price-responsive load shedding we have defined variable $dsm^{S}_{rp,k,i,seg}$, which has an upper bound (3f), appears in the demand balance equation and has a corresponding cost term in the objective function. In order to formulate shifting DSM, we define two variables $dsm^{-}_{rp,k,i,sec}$ and $dsm^{+}_{rp,k,i,sec}$ that represent downward shifting and upward shifting of load. A downward shift represents the amount of power that the load is reduced, so in the power balance this variable can be interpreted as reducing demand by that amount, whereas an upward shift increases demand. In the objective function we consider an explicit cost of a downward shift, which is what the system operator would have to pay a consumer for reducing his/her load. Constraint (3a) defines a balance constraint that ensures that the total upward shifted demand equals the total downward shifted demand over the duration of one representative period. In order to make sure that these shifts are carried within the allowed range of hours, we introduce constraint (3b).

$$\sum_{k} dsm^{+}_{rp,k,i,sec} = \sum_{k} dsm^{-}_{rp,k,i,sec} \quad \forall rp, i, sec \tag{3a}$$

$$dsm^{+}_{rp,k,i,sec} \leq \sum_{kk \in dkk(rp,k,kk,sec)} dsm^{-}_{rp,kk,i,sec} \quad \forall rp, k, i, sec \tag{3b}$$

$$dsm^{+}_{rp,k,i,sec} + dsm^{-}_{rp,k,i,sec} \leq max(\overline{DSM}^{+}, \overline{DSM}^{-}) \quad \forall rp, k, i, sec \tag{3c}$$

$$0 \leq dsm^{-}_{rp,k,i,sec} \leq \overline{DSM}^{-}_{rp,k,i,sec} v^{+/-}_{rp,k,i,sec} \quad \forall rp,k,i,sec \qquad (3d)$$

$$0 \leq dsm^{+}_{rp,k,i,sec} \leq \overline{DSM}^{+}_{rp,k,i,sec}(1 - v^{+/-}_{rp,k,i,sec}) \quad \forall rp,k,i,sec \qquad (3e)$$

$$0 \leq dsm^{S}_{rp,k,i,seg} \leq \overline{DSM}^{S}_{sec} D^{P}_{rp,k,i} \quad \forall rp,k,i,seg \qquad (3f)$$

$$v^{+/-}_{rp,k,i,sec} \in \{0,1\} \qquad (3g)$$

*11.3. Transmission expansion planning (DC + AC)*

LEGO is flexible and can be solved as a single-node problem, a DC optimal power flow (OPF), and a second-order cone programming (SOCP) approximation of the full AC-OPF. In this paper, we have extended LEGO to include also transmission expansion planning (TEP) of candidate lines. If candidate lines are specified in data sheet Network, see Figure 3, LEGO carries out transmission expansion planning, otherwise it is solved as an OPF. In the remainder of this section we describe the mathematical TEP formulation in DC in section 11.3.1, and in AC in section 11.3.2.

*11.3.1. DC-OPF*

Constraints (4) represent the optimal power flow in DC and the additional constraints to account for TEP: active power balance constraint (4a), which contains generation of thermal, renewable and storage technologies, flow to and from the bus, non-supplied energy, and demand side management (shedding and down shifting) and sets it equal to power demand, upward DSM shifting and power consumption of hydrogen units; definition of power flow variable using angle differences for existing (4b) and candidate power lines (4c)-(4d); lower and upper bounds on power flow for existing (4e) and candidate lines (4f); definition of investment variables as binaries by (4g).

$$\sum_{gi(t,i)} p_{rp,k,t} + \sum_{gi(r,i)} p_{rp,k,r} + \sum_{gi(s,i)} (p_{rp,k,s} - cs_{rp,k,s})$$
$$+ \sum_{ijc(j,i,c)} f^{P}_{rp,k,j,i,c} - \sum_{ijc(i,j,c)} f^{P}_{rp,k,i,j,c}$$

$$+pns_{rp,k,i} + \sum_{sec} dsm^{-}_{rp,k,i,sec} + \sum_{seg} dsm^{S}_{rp,k,i,seg}$$

$$= D^{P}_{rp,k,i} + \sum_{sec} dsm^{+}_{rp,k,i,sec} + \sum_{h2gi(h2g,i)} cs^{E}_{rp,k,h2g} \quad \forall rp,k,i \tag{4a}$$

$$f^{P}_{rp,k,ijc} = \frac{(\theta_{rp,k,i} - \theta_{rp,k,j})SB}{Reac_{i,j,c}} \quad \forall rp,k,ijce(i,j,c) \tag{4b}$$

$$f^{P}_{rp,k,ijc} \leq \frac{(\theta_{rp,k,i} - \theta_{rp,k,j})SB}{Reac_{i,j,c}} + \overline{T}_{i,j,c}(1 - x^{L}_{i,j,c}) \quad \forall rp,k,ijcc(i,j,c) \tag{4c}$$

$$f^{P}_{rp,k,ijc} \geq \frac{(\theta_{rp,k,i} - \theta_{rp,k,j})SB}{Reac_{i,j,c}} - \overline{T}_{i,j,c}(1 - x^{L}_{i,j,c}) \quad \forall rp,k,ijcc(i,j,c) \tag{4d}$$

$$-\overline{T}_{i,j,c} \leq f^{P}_{rp,k,i,j,c} \leq \overline{T}_{i,j,c} \quad \forall rp,k,ijce(i,j,c) \tag{4e}$$

$$-\overline{T}_{i,j,c}x^{L}_{i,j,c} \leq f^{P}_{rp,k,i,j,c} \leq \overline{T}_{i,j,c}x^{L}_{i,j,c} \quad \forall rp,k,ijcc(i,j,c) \tag{4f}$$

$$x^{L}_{i,j,c} \in \{0,1\} \quad \forall ijcc(i,j,c) \tag{4g}$$

### *11.3.2. AC-OPF*

In this section we present the mathematical formulation of the power flow equations in AC, which have been approximated as a second-order cone program (SOCP) as in [10] and [11].[12] Note that we have introduced candidate lines in this formulation in order to carry out TEP in an AC-OPF approximation. In LEGO, in the Parameters sheet, one can specify if a DC-OPF or an SOCP approximation of the AC-OPF should be run. The GAMS code itself then replaces (4) with (5) internally.

Constraints (5) represent: an updated version of: the active power balance equation (5a); a reactive power balance (5b); definition of active power flow from bus $i$ to bus $j$ (5c) and from bus $j$ to bus $i$ (5d) for existing lines[13];

[12]The following auxiliary variables have been used in the SOCP formulation: $cii_{rp,k,i}, cij_{rp,k,i,j}, sij_{rp,k,i,m}$, which represent the square of the voltage at bus $i$, the product of the voltages at bus $i$ and $j$ times $cos(\theta_{ij})$, and the product of the voltages at bus $i$ and $j$ times $sin(\theta_{ij})$.

[13]In the DC-OPF we do not explicitly consider both directions of the power flow, demonstrated by the fact that equation (4b) is only defined in one direction. This is connected to the underlying DC hypothesis that $f^{P}_{i,j} = -f^{P}_{i,j}$, which is not necessarily the case in actual AC power flow. That is why in the SOCP formulation we have to explicitly define and distinguish

definition of active power flow from bus $i$ to bus $j$ (5e) and from bus $j$ to bus $i$ (5f) for candidate lines; definition of reactive power flow from bus $i$ to bus $j$ (5g) and from bus $j$ to bus $i$ (5h) for existing lines; definition of reactive power flow from bus $i$ to bus $j$ (5i) and from bus $j$ to bus $i$ (5j) for candidate lines; the conic constraint of the SOCP representing the squares of voltages (5k); ensuring maximum angle differences (5l); lower and upper bounds on reactive power provided by FACTS (5m); lower and upper bounds on reactive power provided by thermal units (5n); lower and upper bounds on reactive power provided by unit $g$ (5o)[14]; bounds on active (5p) and reactive (5q) power flow for existing lines; bounds on active (5r) and reactive (5s) power flow for candidate lines; bounds of auxiliary $cii$ variable (5t); bounds of auxiliary $cij$ variable (5u); bounds of auxiliary $sij$ variable (5v).

$$\begin{aligned}
\sum_{gi(t,i)} p_{rp,k,t} + \sum_{gi(r,i)} p_{rp,k,r} + \sum_{gi(s,i)} p_{rp,k,s} - \sum_{gi(s,i)} cs_{rp,k,s} + pns_{rp,k,i} \\
+ \sum_{sec} dsm^{-}_{rp,k,i,sec} + \sum_{seg} dsm^{S}_{rp,k,i,seg} \\
= \sum_{(j,c)\in ijc(i,j,c)} f^{P}_{rp,k,i,j,c} + \sum_{(j,c)\in ijc(j,i,c)} f^{P}_{rp,k,i,j,c} \\
+ cii_{rp,k,i} G_i SB + D^{P}_{rp,k,i} \\
+ \sum_{sec} dsm^{+}_{rp,k,i,sec} + \sum_{h2gi(h2g,i)} cs^{E}_{rp,k,h2g} \quad \forall rp,k,i
\end{aligned} \tag{5a}$$

$$\begin{aligned}
\sum_{gi(t,i)} q_{rp,k,t} + \sum_{gi(r,i)} q_{rp,k,i} + \sum_{gi(s,i)} q_{rp,k,s} + \sum_{gi(facts,i)} q_{rp,k,facts} \\
+ pns_{rp,k,i} R_i + \sum_{sec} dsm^{-}_{rp,k,i,sec} R_i + \sum_{seg} dsm^{S}_{rp,k,i,seg} R_i \\
= \sum_{(j,c)\in ijc(i,j,c)} f^{Q}_{rp,k,i,j,c} + \sum_{(j,c)\in ijc(j,i,c)} f^{Q}_{rp,k,i,j,c} \\
- cii_{rp,k,i} B_i SB + D^{Q}_{rp,k,i}
\end{aligned}$$

between the direction of the power flow.

[14] This constraint is redundant for thermal and FACTS given that there are more binding constraints for them available. This is just a general limit to help the numerical solvers, and for storage and renewable units should they be able to provide reactive power.

$$+\sum_{sec} dsm^{+}_{rp,k,i,sec} R_i + \sum_{h2gi(h2g,i)} cs^{E}_{rp,k,h2g} R_i \quad \forall rp,k,i \tag{5b}$$

$$f^{P}_{rp,k,i,j,c} = SB[G_{i,j,c} cii_{rp,k,i} - cij_{rp,k,i,j} G_{i,j,c} + sij_{rp,k,i,j} B_{i,j,c}]$$
$$\forall rp,k,ijce(i,j,c) \tag{5c}$$

$$f^{P}_{rp,k,j,i,c} = SB[G_{i,j,c} cii_{rp,k,j} - cij_{rp,k,i,j} G_{i,j,c} - sij_{rp,k,i,j} B_{i,j,c}]$$
$$\forall rp,k,ijce(i,j,c) \tag{5d}$$

$$SB[G_{i,j,c} cii_{rp,k,i} - cij_{rp,k,i,j} G_{i,j,c} + sij_{rp,k,i,j} B_{i,j,c}] - M^{L}(1 - x^{L}_{i,j,c})$$
$$\leq f^{P}_{rp,k,i,j,c} \leq$$
$$SB[G_{i,j,c} cii_{rp,k,i} - cij_{rp,k,i,j} G_{i,j,c} + sij_{rp,k,i,j} B_{i,j,c}] + M^{L}(1 - x^{L}_{i,j,c})$$
$$\forall rp,k,ijcc(i,j,c) \tag{5e}$$

$$SB[G_{i,j,c} cii_{rp,k,j} - cij_{rp,k,i,j} G_{i,j,c} - sij_{rp,k,i,j} B_{i,j,c}] - M^{L}(1 - x^{L}_{i,j,c})$$
$$\leq f^{P}_{rp,k,j,i,c} \leq$$
$$SB[G_{i,j,c} cii_{rp,k,j} - cij_{rp,k,i,j} G_{i,j,c} - sij_{rp,k,i,j} B_{i,j,c}] + M^{L}(1 - x^{L}_{i,j,c})$$
$$\forall rp,k,ijcc(i,j,c) \tag{5f}$$

$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,i} + sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$= f^{Q}_{rp,k,i,j,c} \quad \forall rp,k,ijce(i,j,c) \tag{5g}$$

$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,j} - sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$= f^{Q}_{rp,k,j,i,c} \quad \forall rp,k,ijce(i,j,c) \tag{5h}$$

$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,i} + sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$-M^{L}(1 - x^{L}_{i,j,c}) \leq f^{Q}_{rp,k,i,j,c} \leq M^{L}(1 - x^{L}_{i,j,c}) +$$
$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,i} + sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$\forall rp,k,ijcc(i,j,c) \tag{5i}$$

$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,j} - sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$-M^{L}(1 - x^{L}_{i,j,c}) \leq f^{Q}_{rp,k,j,i,c} \leq M^{L}(1 - x^{L}_{i,j,c}) +$$
$$SB[-(B_{i,j,c} + Bc_{i,j,c}/2) cii_{rp,k,j} - sij_{rp,k,i,j} G_{i,j,c} + cij_{rp,k,i,j} B_{i,j,c}]$$
$$\forall rp,k,ijcc(i,j,c) \tag{5j}$$

$$cij^{2}_{rp,k,i,j} + sij^{2}_{rp,k,i,j} \leq cii_{rp,k,i} cii_{rp,k,j} \quad \forall rp,k,line(i,j) \tag{5k}$$

$$-cij_{rp,k,i,j}tan(\Delta) \leq sij_{rp,k,i,j} \leq cij_{rp,k,i,j}tan(\Delta) \quad \forall rp,k,line(i,j) \tag{5l}$$

$$x_{facts}\underline{Q}_{facts} \leq q_{rp,k,facts} \leq x_{facts}\overline{Q}_{facts} \quad \forall rp,k,facts \tag{5m}$$

$$u_{rp,k,t}\underline{Q}_t \leq q_{rp,k,t} \leq u_{rp,k,t}\overline{Q}_t \quad \forall rp,k,t \tag{5n}$$

$$\underline{Q}_g \leq q_{rp,k,g} \leq \overline{Q}_g \quad \forall rp,k,g \tag{5o}$$

$$-\overline{T}_{i,j,c} \leq f^P_{rp,k,i,j,c} \leq \overline{T}_{i,j,c} \quad \forall rp,k,ijce(i,j,c) \tag{5p}$$

$$-\overline{A}_{i,j,c} \leq f^Q_{rp,k,i,j,c} \leq \overline{A}_{i,j,c} \quad \forall rp,k,ijce(i,j,c) \tag{5q}$$

$$-\overline{T}_{i,j,c}x^L_{i,j,c} \leq f^P_{rp,k,i,j,c} \leq \overline{T}_{i,j,c}x^L_{i,j,c} \quad \forall rp,k,ijcc(i,j,c) \tag{5r}$$

$$-\overline{A}_{i,j,c}x^L_{i,j,c} \leq f^Q_{rp,k,i,j,c} \leq \overline{A}_{i,j,c}x^L_{i,j,c} \quad \forall rp,k,ijcc(i,j,c) \tag{5s}$$

$$\underline{V}_i^2 \leq cii_{rp,k,i} \leq \overline{V}_i^2 \quad \forall rp,k,i \tag{5t}$$

$$\underline{V}_i^2 \leq cij_{rp,k,i,j} \leq \overline{V}_i^2 \quad \forall rp,k,line(i,j) \tag{5u}$$

$$-\overline{V}_i^2 \leq sij_{rp,k,i,j} \leq \overline{V}_i^2 \quad \forall rp,k,line(i,j) \tag{5v}$$

### *11.4. Hydrogen*

In LEGO we consider the possibility of producing hydrogen via electrolysis, powered by energy, in order to fulfill an inelastic demand of hydrogen. To that purpose, we define an index $h2g$, which are hydrogen generating units and in this particular case that would correspond to electrolyzers, each of which consume a particular amount of electric energy $cs^E_{rp,k,h2g}$ (MW) in order to produce hydrogen $p^{H2}_{rp,k,h2g}$ (kg), whose bounds are given in (6a). We employ conversion factor $HPE_{h2g}$, which quantifies how many kg of hydrogen are obtained through electrolysis when using 1 MWh of electric energy, given in (6b). Note that a-priori, hydrogen might flow in a network (i.e., pipelines) different from the transmission grid. Hence, we define index $h2i$, which corresponds to the different nodes of the hydrogen network. However, since an electrolyzer does consume electric energy, it will likely be located at a node $h2i$ that actually coincides bus $i$ of the transmission network. In any case, dynamic set $h2gi(h2g,i)$ determines at what bus the unit $h2g$ is located in order to add the corresponding consumption of electric energy in the power balance equation given in section 11.3. There could be different sectors $h2sec$ in which we have a demand for hydrogen. Constraint

(6c) represents the hydrogen balance equation for each sector, which contains the hydrogen production $p^{H2}_{rp,k,h2g}$, and the hydrogen non-served $h2ns_{rp,k,h2i}$ (6d). Constraint (6e) defines discrete investment variables. In future work, we plan to expand the formulation of the hydrogen sector in order to contain a network of pipelines, and other hydrogen-related infrastructure.

$$0 \leq p^{H2}_{rp,k,h2g} \leq \overline{P}^{E}_{h2g} W^{K}_{k} HPE_{h2g}(x^{H2}_{h2g} + EU^{H2}_{h2g}) \quad \forall rp, k, h2g \quad \text{(6a)}$$

$$cs^{E}_{rp,k,h2g} W^{K}_{k} HPE_{h2g} = p^{H2}_{rp,k,h2g} \quad \forall rp, k, h2g \quad \text{(6b)}$$

$$\sum_{h2gh2i(h2g,h2i)} p^{H2}_{rp,k,h2g} + h2ns_{rp,k,h2i} = \sum_{h2sec} D^{H2}_{rp,k,h2i,h2sec} \quad \forall rp, k, h2i \quad \text{(6c)}$$

$$0 \leq h2ns_{rp,k,h2i} \leq \sum_{h2sec} D^{H2}_{rp,k,h2i,h2sec} \quad \forall rp, k, hi \quad \text{(6d)}$$

$$x^{H2}_{h2g} \in \mathbb{Z}^{+,0}, x^{H2}_{h2g} \leq \overline{X}^{H2}_{h2g} \quad \forall h2g \quad \text{(6e)}$$

*11.5. Standard Constraints*

Constraints (7) contain all constraints regarding thermal generators: upward reserve requirement (7a); downward reserve requirement (7b); definition of total power output with the technical minimum and output above the technical minimum (7c); limit of upward reserve in case start-up occurred (7d); limit of upward reserve in case shut-down occurs (7e); limit of downward reserve (7f); definition of commitment, start-up and shut-down logic (7g); upper bound of commitment variable (7h); ramp-up constraint (7i); ramp-down constraint (7j); lower and upper bound of total power output (7k); lower and upper bound of reserves and output above the minimum (7l); definition of logical variables as binaries (7m).

With this in mind, we quickly want to define the notation of double minus $--$ or double plus $++$ that appears sometimes in the remainder of this section. The term $k{-}{-}1$ simply refers to the previous within-time period $k$. For example, if $k = 2$, then $k{-}{-}1$ corresponds to $k = 1$. But, if $k = 1$, then $k{-}{-}1$ corresponds to $k = 24$. The double minus creates a cyclic link between the first and the last

$k$ of the same representative period. In the remainder of the paper, we use this terminology for commitment variables and for cyclic storage constraints.

$$\sum_t res^+_{rp,k,t} + \sum_s res^+_{rp,k,s} \geq RES^+ \sum_i D^P_{rp,k,i} \quad \forall rp, k \tag{7a}$$

$$\sum_t res^-_{rp,k,t} + \sum_s res^-_{rp,k,s} \geq RES^- \sum_i D^P_{rp,k,i} \quad \forall rp, k \tag{7b}$$

$$p_{rp,k,t} = u_{rp,k,t}\underline{P}_t + \hat{p}_{rp,k,t} \quad \forall rp, k, t \tag{7c}$$

$$\hat{p}_{rp,k,t} + res^+_{rp,k,t} \leq (\overline{P}_t - \underline{P}_t)(u_{rp,k,t} - y_{rp,k,t}) \quad \forall rp, k, t \tag{7d}$$

$$\hat{p}_{rp,k,t} + res^+_{rp,k,t} \leq (\overline{P}_t - \underline{P}_t)(u_{rp,k,t} - z_{rp,k++1,t}) \quad \forall rp, k, t \tag{7e}$$

$$\hat{p}_{rp,k,t} \geq res^-_{rp,k,t} \quad \forall rp, k, t \tag{7f}$$

$$u_{rp,k,t} - u_{rp,k--1,t} = y_{rp,k,t} - z_{rp,k,t} \quad \forall rp, k, t \tag{7g}$$

$$u_{rp,k,t} \leq x_t + EU_t \quad \forall rp, k, t \tag{7h}$$

$$\hat{p}_{rp,k,t} - \hat{p}_{rp,k--1,t} + res^+_{rp,k,t} \leq u_{rp,k,t}RU_t \quad \forall rp, k, t \tag{7i}$$

$$\hat{p}_{rp,k,t} - \hat{p}_{rp,k--1,t} - res^-_{rp,k,t} \geq -u_{rp,k--1,t}RD_t \quad \forall rp, k, t \tag{7j}$$

$$0 \leq p_{rp,k,t} \leq \overline{P}_t(x_t + EU_t) \quad \forall rp, k, t \tag{7k}$$

$$0 \leq \hat{p}_{rp,k,t}, res^-_{rp,k,t}, res^+_{rp,k,t} \leq (\overline{P}_t - \underline{P}_t)(x_t + EU_t) \quad \forall rp, k, t \tag{7l}$$

$$u_{rp,k,t}, y_{rp,k,t}, z_{rp,k,t} \in \{0, 1\} \quad \forall rp, k, t \tag{7m}$$

Constraint (8a) represents the inter-period evolution of the storage state of charge; upper bound of inter storage state of charge (8b); lower bound of inter storage state of charge (8c); cyclic storage constraint (8d); intra-period evolution of storage state of charge (8e); bound of upward reserve (8f); bound of downward reserve (8g); upper bound of intra storage state of charge (8h); lower bound of intra storage state of charge (8i); to avoid simultaneous charging and discharging (8j); definition of binary variable to avoid simultaneous charging and discharging (8k); lower and upper bounds on production, consumption and reserve variables (8l); lower and upper bound of intra storage state of charge (8m); lower and upper bound on spillages (8n).

$$inter_{p,s} = inter_{p-MOW,s} + InRes_{s,p=MOW}$$
$$+ \sum_{\Gamma(p-MOW \leq pp \leq p,rp,k)} (-sp_{rp,k,s} + IF_{rp,k,s} W_k^K$$
$$-p_{rp,k,s} W_k^K / \eta_s^{DIS} + cs_{rp,k,s} W_k^K \eta_s^{CH}) \quad \forall p, s \tag{8a}$$
$$inter_{p,s} \leq \overline{P}_s ETP_s (x_s + EU_s) \quad \forall s, p : mod(p, MOW) = 0 \tag{8b}$$
$$inter_{p,s} \geq \underline{R}_s \overline{P}_s ETP_s (x_s + EU_s) \quad \forall s, p : mod(p, MOW) = 0 \tag{8c}$$
$$inter_{p,s} \geq InRes_{s,p} \quad \forall s, p = CARD(p) \tag{8d}$$
$$intra_{rp,k,s} = intra_{rp,k--1,s} - sp_{rp,k,s} + IF_{rp,k,s} W_k^K$$
$$-p_{rp,k,s} W_k^K / \eta_s^{DIS} + cs_{rp,k,s} W_k^K \eta_s^{CH} \quad \forall rp, k, s \tag{8e}$$
$$\hat{p}_{rp,k,s} - cs_{rp,k,s} + res^+_{rp,k,s} \leq \overline{P}_s (bx_s + EU_s) \quad \forall rp, k, s \tag{8f}$$
$$\hat{p}_{rp,k,s} - cs_{rp,k,s} - res^-_{rp,k,s} \geq -\overline{P}_s (bx_s + EU_s) \quad \forall rp, k, s \tag{8g}$$
$$intra_{rp,k,s} \leq \overline{P}_s ETP_s (x_s + EU_s)$$
$$-(res^-_{rp,k,s} + res^-_{rp,k--1,s}) W_k^K \quad \forall rp, k, s \tag{8h}$$
$$intra_{rp,k,s} \geq \underline{R}_s \overline{P}_s ETP_s (x_s + EU_s)$$
$$+(res^+_{rp,k,s} + res^+_{rp,k--1,s}) W_k^K \quad \forall rp, k, s \tag{8i}$$
$$p_{rp,k,s} \leq b^{ch/d}_{rp,k,s} M^{ch/d}, cs_{rp,k,s} \leq (1 - b^{ch/d}_{rp,k,s}) M^{ch/d} \quad \forall rp, k, s \tag{8j}$$
$$b^{ch/d}_{rp,k,s} \in \{0, 1\} \quad \forall rp, k, s \tag{8k}$$
$$0 \leq p_{rp,k,s}, cs_{rp,k,s} res^-_{rp,k,s}, res^+_{rp,k,s} \leq \overline{P}_s (bx_s + EU_s) \quad \forall rp, k, s \tag{8l}$$
$$ETP_s \overline{P}_s \underline{R}_s (x_s + EU_s) \leq intra_{rp,k,s} \leq ETP_s \overline{P}_s (x_s + EU_s) \quad \forall rp, k, s \tag{8m}$$
$$0 \leq sp_{rp,k,s} \leq (1 - \overline{R}_s) ETP_s \overline{P}_s (x_s + EU_s) \quad \forall rp, k, s = hydro \tag{8n}$$

*11.6. Battery degradation*

The constraints in (8) do not account for battery degradation and the corresponding arising costs. According to [], battery degradation has a significant impact on both battery operation and investment decisions. Therefore, we have decided to include the option of considering battery degradation in our model, following the methodology proposed by [**?** ] that approximates the rainflow algorithm, as follows in (9). To that purpose we introduce the additional index $a$ that represents the segments of the cycle aging cost function. If battery degradation is considered, then constraints (9) are included in the modeling process, and (8e) is no longer written for storage technologies $s$ that represent batteries.

$$p_{rp,k,s} = \sum_a dis_{rp,k,s,a} \quad \forall rp, k, cdsf(s) \tag{9a}$$

$$cs_{rp,k,s} = \sum_a ch_{rp,k,s,a} \quad \forall rp, k, cdsf(s) \tag{9b}$$

$$intra_{rp,k,s} = \sum_a soc_{rp,k,s,a} \quad \forall rp, k, cdsf(s) \tag{9c}$$

$$0 \leq soc_{rp,k,s,a} \leq \overline{P}_s(x_s + EU_s)ETP_s/CARD(a) \quad \forall rp, k, cdsf(s), a \tag{9d}$$

$$\begin{aligned} soc_{rp,k,s,a} = soc_{rp,k--1,s,a} - dis_{rp,k,s,a} W_k^k / \eta_s^{DIS} \\ + ch_{rp,k,s,a} W_k^k \eta_s^{CH} \quad \forall rp, k, cdsf(s) \end{aligned} \tag{9e}$$

*11.7. Modeling Inertia*

Let us now describe in detail the corresponding inertia constraints (10) for a generation expansion framework: definition of the scaled power gain factor of a thermal unit (10a), which represents how much one particular (dispatched) thermal unit can contribute with respect to the total dispatched thermal capacity in the system; definition of the scaled power gain factor of a virtual unit (such as a wind turbine or a battery) (10b), which is defined slightly differently from the corresponding factor of a synchronous generator because it does not have a commitment variable. It is therefore defined as the fraction of its current power output over the total available virtual power output in this moment[15]. Definition of inertia provided by synchronous generators (10c) and virtual generators (10d). The right-hand side of (10d) is multiplied by $x_v$ because $k_{rp,k,v}$ represents the power gain factor for one unit of the virtual technology $v$. However, in total $x_v$ total units of technology $v$ are built. In (10c) this number is

[15]In a system where there is a large amount of wind curtailment, this would be an overly conservative approximation.

guaranteed to be 1, and that is why it is not explicitly modeled in (10c). Definition of total system inertia (10e): note that total system inertia is a weighted average of virtual inertia and synchronous generator inertia. Rate of change of frequency (RoCoF) constraint (10f). Lower and upper bounds on inertia variables (10g). Lower and upper bounds on power gain factors (10h). The nonlinear constraints in (10), and in particular constraints (10a), (10b), (10d), and (10e), are linearized as described in [8].

$$k_{rp,k,t} = \frac{\overline{P}_t}{\sum_{tt} \overline{P}_{tt} u_{rp,k,tt}} u_{rp,k,t} \quad \forall rp, k, t \tag{10a}$$

$$k_{rp,k,v} = \frac{p_{rp,k,t}}{\sum_{vv} \overline{P}_{vv}(x_{vv} + EU_{vv}) PF_{rp,k,vv}} \quad \forall rp, k, v \tag{10b}$$

$$M^{SG}_{rp,k} = \sum_t 2k_{rp,k,t} H_t \quad \forall rp, k \tag{10c}$$

$$M^{VI}_{rp,k} = \sum_v 2k_{rp,k,v} H_v x_v \quad \forall rp, k \tag{10d}$$

$$M_{rp,k} = \frac{M^{SG}_{rp,k} \sum_{tt} \overline{P}_{tt} u_{rp,k,tt} + M^{VI}_{rp,k} \sum_{vv} \overline{P}_{vv}(x_{vv} + EU_{vv}) PF_{rp,k,vv}}{\sum_{tt} \overline{P}_{tt} u_{rp,k,tt} + \sum_{vv} \overline{P}_{vv}(x_{vv} + EU_{vv}) PF_{rp,k,vv}} \quad \forall rp, k \tag{10e}$$

$$\frac{\dot{f}_{lim}}{f_b} M_{rp,k} \geq \Delta P_{rp,k} \quad \forall rp, k \tag{10f}$$

$$0 \leq M^{VI}_{rp,k}, M^{SG}_{rp,k}, M_{rp,k} \leq \overline{M} \quad \forall rp, k \tag{10g}$$

$$0 \leq k_{rp,k,g} \leq 1 \quad \forall rp, k, g \tag{10h}$$

## 12. APPENDIX C: Literature Review

Let us now compare LEGO to some other available open-source ESMs. The GenX[16] model [12] originally developed at MIT is a powerful electricity resource capacity expansion planning tool. Some differences with respect to LEGO are the fact that GenX does not account for RoCoF constraints, and currently considers a zonal flow. A DC power flow is in development but there

[16]https://github.com/GenXProject/GenX

is not the option of considering an AC power flow. With respect to temporal flexibility, GenX works for chronological time periods (including representative days). LEGO can be run for both chronological and non-chronological time periods, and is capable of establishing inter-temporal constraints [13] between non-chronological representative periods, which allows combining short- and long-term storage in aggregated optimization models. There are not many aggregated models that incorporate long-term storage as pointed out in [14]. The Electricity Market Model EMMA[17] [15] is a techno-economic dispatch and investment model developed by NEON consulting company. EMMA does not account for hydro reservoir modeling, does not include DSM, RoCoF or UC constraints, nor does it allow for discrete investment decisions. It also does not model an optimal power flow or the hydrogen sector. The openTEPES[18] model [16] developed at the Institute for Research in Technology of Comillas Pontifical University is a cost minimization GEPTEP model. Its time representation allows for chronological periods (of one or multiple hours); however, it does not allow for representation of non-chronological representative periods in combination with long-term storage technologies as LEGO does, nor does it consider an AC-OPF, RoCoF, DSM nor the hydrogen sector. The Next Energy Modeling system for Optimization (NEMO)[19] is an open-source energy system optimization tool developed by the Stockholm Environment Institute. NEMO does not allow for AC-OPF, nor does it consider RoCoF, DSM or hydrogen. The temporal structure uses (hierarchical) time slices, which need to be ordered in order to account for storage. The ANTARES[20] tool for adequacy reporting of electric systems [17], is an open source power system simulator to quantify the adequacy or the economic performance of interconnected energy systems. The underlying optimization models of ANTARES work with chronological time periods (hours within weeks). In order to incorporate long-term storage technologies, such as

[17] https://neon.energy/en/emma/
[18] https://github.com/IIT-EnergySystemModels/openTEPES
[19] https://github.com/sei-international/NemoMod.jl
[20] https://github.com/AntaresSimulatorTeam/Antares_Simulator

hydro, additional information, such as the current weekly water value, has to be incorporated as data. The temporal structure of LEGO is more flexible than that and does therefore not rely on exogenous data. Moreover, ANTARES does not incorporate DSM, the hydrogen sector nor an AC-OPF formulation.

The PyPSA [18] framework seems to run on chronological time periods. They can be weighted in order to represent durations of more than one hour for example (and these weights can also represent probability). Moreover, they do account for investment periods, which is something that LEGO does not currently have. However, from what we understand, this is not the typical framework for representative days (because time periods are linked, which does not necessarily have to be the case with representative days) and it might be difficult to simultaneously optimize both short- and long-term storage technologies. To the best of our knowledge PyPSA is capable of carrying out a "simulation" of the full AC power flow. So indeed, it does support non-linear power flow for AC networks, but this is not used for and in the optimization process. In the optimization process, PyPSA uses a DC-OPF approximation for the AC power flow. LEGO does not carry out power flow simulations, it is an optimization (not simulation model) and it allows for optimizing both the DC and the (SOCP approximation of an) AC power flow. The Switch [19] framework also has investment periods and it also seems to have chronological time periods that can have variable duration. In the Switch framework it seems that only a single-node transport model is employed. The SpineOpt [20] framework is, in our opinion, the most advanced with respect to the temporal flexibility. They have temporal blocks (e.g. representative days) that can be disconnected and optimized. They allow for different time scales for different sectors, which is also something that LEGO does not allow currently. If we understand the documentation correctly, SpineOpt also seems to have a similar inter-day storage treatment as LEGO. However, SpineOpt uses a DC-OPF formulation using PTDFs and does not allow for an AC-OPF framework.

DSM is gaining attention and importance in power systems, especially with respect to the integration of variable renewable energy sources [21] or otherwise improving system efficiency [22]. In LEGO, we wanted to include two types of DSM: one that was simple load shedding at a previously established cost; and another, that was load shifting in order to be able to accommodate potentials of DSM via heating and cooling, electric vehicles (EVs) etc. There are several approaches in the literature to address shifting DSM. For example, [23] presents a DSM formulation that allows for load shifting but does not guarantee to avoid simultaneous up and downward shifting. In [24], the authors propose an alternative version of [23] that tries to circumvent this problem, but in order to do so, they have to introduce a downward DSM variable that depends on the temporal index twice. This means that the total number of downward DSM variables is the cardinality of the time index squared. An approach like this works when assessing one single representative day of 24 consecutive hourly time periods, but not when assessing long-term problems, where 8760 consecutive hours are analyzed. Based on [24], our DSM variables only depend on the time index, once, and we avoid simultaneous up and downward DSM by introducing binary variables.

With respect to hydrogen, currently, only few GEPTEP models consider hydrogen technologies and even fewer are open-source. As an example, the authors in [25] model a hydrogen technology chain formed by electrolyzer – methanization – storage tanks – OCGT – CCUS with symmetric charging and discharging capacity. However, their model is formulated as an LP which is not capable of considering discrete investment decisions. The comprehensive model presented in [12] considers a combination of electroyzer – hydrogen storage – fuel cell with asymmetric charging and discharging capacity but they only apply a simple transport problem to model inter-zonal power flows. Authors in [26] take a game-theoretic approach to modeling integrated planning of power and natural gas systems in a cooperative environment. They incorporate Power-to-Gas units with the option of providing up and downward reserve capacity. However, their model is formulated as an MINLP which is highly non-convex and non-linear. The authors in [27] propose a multi-objective model. They minimize the average costs of carbon emission reduction and maximize profits of Power-to-Gas units in order to determine their optimal capacity and grid connection points. Although their model focuses on transformation of integrated power and gas systems they do not explicitly consider battery technologies.

LEGO is based on preliminary work [8] by the authors, and has been extended to incorporate many novel aspects, such as transmission expansion planning in both a DC- and AC framework, the introduction of DSM, the hydrogen sector, battery degradation, the economic output based on dual variables, and the fact that it is now freely available on GitHub[21]. To the best of our knowledge, there is no single open-source model available that combines all of the previously mentioned characteristics, which makes LEGO uniquely versatile.

## 13. APPENDIX D: Case Studies

The case studies presented in this section are designed to showcase the modularity and flexibility of LEGO. Section 13.1 compares an exact hourly generation expansion planning DC-OPF with its corresponding model using only 7 representative days. Section 13.2 considers DSM and section 13.3 assesses hydrogen production, two very promising future technologies that could shift optimal power system planning and operation. Finally, in section 13.4 we carry out a transmission (TEP) and generation (GEP) expansion planning problem that contrasts optimal decision making in an AC framework as opposed to a DC simplification.

### *13.1. Hourly model versus representative days model*

The main objective of this section is to show that the LEGO model supports exact hourly data, as well as approximations of that data via representative days. Instead of having to commit to one inflexible data structure that binds the modeler to a particular time representation, LEGO is capable of handling both, which allows the modeler to decide (or even change ex-post depending on the application).

[21]https://github.com/IEE-TUGraz/LEGO

To that purpose we run an rMIP version[22] of LEGO for the exact hourly data set over the time horizon of an entire year, and for an approximated model using 7 representative days that have been chosen applying k-means to the full time series. Please note that clustered model results depend heavily on the clustering technique; however, it is beyond the scope of this paper to assess the best clustering techniques for expansion planning. We simply picked k-means because it is a standard clustering technique.

Moreover, in this case we do not enforce a particular renewable penetration. However, we do require a 110% firm capacity constraint to be met. The model considers an almost greenfield approach. The only existing technology in the system is hydro. The reason for this is twofold: first, when assessing economic results we wanted to show the difference between existing units and new investments; second, on the transition towards a carbon-neutral future, existing thermal generation assets might be retired, but hydro infrastructure will definitely remain as long as the useful life allows. All other technologies[23] are considered candidate units for investment in a DC-OPF setting. Note that in this case study, there is no transmission expansion.

The values shown in Table 52 represent capacity investment decisions per technology for the representative days case study and the hourly rMIP models. The results show very clearly the importance of accounting for hourly details in generation expansion planning. When considering only 7 representative days, the necessary capacity in peakers, such as OCGTs, is under-estimated (1427 versus 2240 MW). It also seems that when accounting for all individual hours of the year, about 300 MW more CCGT capacity is required. On the other hand, the representative day model version over-estimates the optimal wind capacity

[22] The hourly model consists of 8760 individual hours. Running a MIP expansion and UC model for the full hourly model renders models that would exceed the computational infrastructure available to us. Hence, we have decided to run this section as rMIP model for a fair comparison. Note that in other section we do run a MIP version of the representative day model as well.

[23] Coal, gas-fired units, solar, wind and batteries.

by almost 600 MW. The results show to what extent the representative days under-estimate the uncertainty caused by renewable energy sources, which - in reality - would have to be compensated by thermal power plants.

Differences between full hourly and representative days models are to be expected and could be remedied by either increasing the number of representative days, or by choosing representative days more intelligently (than just doing k-means), for example by including extreme-value days.

| | **CCGT** | **OCGT** | **BESS** | **Wind** | **Solar** |
|---|---|---|---|---|---|
| 7 LRP | 2738 | 1427 | 210 | 2202 | 1269 |
| Hourly | 3079 | 2240 | 220 | 1673 | 1305 |

Table 52: Capacity investments (MW) per technology in representative days and hourly case studies (firm capacity = 110%)

Table 53 contains the annual production (in GWh) per technology for both temporal representations. In this case study (where no specific renewable penetration is enforced), the main difference in production between the representative days and the full hourly model are 1000 GWh that are shifted from wind to CCGT production. Note that hydro remains the same in both cases, because the maximum amount of hydro production is limited.

| | **CCGT** | **OCGT** | **BESS** | **Wind** | **Solar** | **Hydro** |
|---|---|---|---|---|---|---|
| 7 LRP | 22426 | 130 | 239 | 5032 | 2488 | 1545 |
| Hourly | 23236 | 216 | 277 | 4069 | 2559 | 1545 |

Table 53: Annual production (GWh) per technology in representative days and hourly case studies (firm capacity = 110%)

Finally, in Table 54 we present the economic results obtained through LEGO. Note that LEGO actually yields those results per generating unit, however, for the sake of legibility in Table 54 we have summed them per technology. As discussed in an earlier section, total profits are calculated as: spot market revenues minus spot market purchases (only for BESS), reserve market revenues minus costs, minus O&M costs, minus investment costs, plus RES quota payments (that do not apply here because we are not forcing a particular renewable penetration so they have been omitted in the table), plus firm capacity payments. We can confirm that with a convex model (an rMIP) economic theory is confirmed and new investments fully recover investment costs, as can be seen by the zero profits of all newly built technologies. Hydro was an existing technology and is therefore the only technology that has positive profits in such a model setup.

Table 54 puts in contrast the different economic concepts for the representative days and the full hourly model. This allows us to analyze where the largest deviations between the hourly and the representative days case are. But the profits results in general, are very useful to analyze, for example, in what market a new technology is going to make the most of its money. In the hourly model, the BESS technology makes 1.31 M€ (1.72-0.41) on the reserve market and 8.63 M€ (17.41-8.78) on the spot market. Reserve market profits would be 15% of spot market profits for BESS. For other technologies, this percentage is negligible. These results showcase the contribution of BESS to the reserve market in this case study. We also observe that all technologies obtain firm capacity payments according to the firm capacity installed to make them whole. Note that the model itself determines the firm capacity payment, as the dual variable of the firm capacity constraint. In this case study here, the firm capacity payment amounts to 25.8 k€/MW of firm capacity installed.

### *13.2. Impact of demand-side management*

In this section we demonstrate how the impacts of demand-side management (DSM) can affect power system operations. To that purpose we run the MIP version of LEGO for a 100% renewable penetration (not allowing for thermal investments) with and without activating the DSM option associated to constraints (3).

DSM provides the power system with more flexibility, and it is therefore not surprising that the system cost is 3.7% lower with DSM (2432.2 M€) than

| | CCGT | OCGT | BESS | Wind | Solar | Hydro |
|---|---|---|---|---|---|---|
| **Spot market revenues** | 1086.44 (1135.70) | 9.83 (16.19) | 16.12 (17.41) | 155.81 (118.88) | 106.37 (109.39) | 95.06 (94.89) |
| **Spot market costs** | 0 (0) | 0 (0) | -8.32 (-8.78) | 0 (0) | 0 (0) | 0 (0) |
| **Reserve market revenues** | 0.22 (0.44) | 0.67 (0.84) | 2.06 (1.72) | 0 (0) | 0 (0) | 0.97 (0.66) |
| **Reserve market costs** | -0.22 (-0.44) | -0.26 (-0.35) | -0.49 (-0.41) | 0 (0) | 0 (0) | 0 (0) |
| **O&M costs** | -1039.79 (-1083.25) | -10.24 (-16.68) | -0.95 (-1.11) | -10.06 (-8.14) | 0 (0) | 0 (0) |
| **Investment costs** | -114.51 (-128.75) | -35.37 (-55.51) | -13.61 (-14.29) | -159.96 (-121.54) | -107.19 (-110.23) | 0 (0) |
| **Firm capacity payments** | 67.86 (76.30) | 35.37 (55.51) | 5.19 (5.45) | 14.21 (10.80) | 0.82 (0.84) | 3.87 (3.87) |
| **Total profits** | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 99.90 (99.42) |

Table 54: Annual profit (M€) concepts per technology for representative days and the full hourly case (results in parentheses) for a 110% firm capacity requirement.

without (2525.5 M€). Table 55 contains the changes in capacity investments. Batteries are most affected by the incorporation of DSM and suffer a 17.7% decrease in capacity, whereas wind and solar capacity does not change much. Energy production behaves similarly.

| | **BESS** | **Wind** | **Solar** |
|---|---|---|---|
| Investment w/o DSM (GW) | 9.6 | 3.5 | 18.8 |
| Investment with DSM (GW) | 7.9 (-17.7%) | 3.4 (-2.9%) | 18.6 (-1.1%) |
| Production w/o DSM (GWh) | 11.7 | 4.4 | 26.9 |
| Production with DSM (GWh) | 10.7 (-8.4%) | 4.7 (+5.6%) | 26.5 (-1.3%) |

Table 55: Investments and production per technology with and without DSM in 100% renewable system. Relative change in parentheses.

In Figure 6 we show the behavior of up- and downward DSM for one representative day at bus 6, which is the highest demand bus. In order to understand

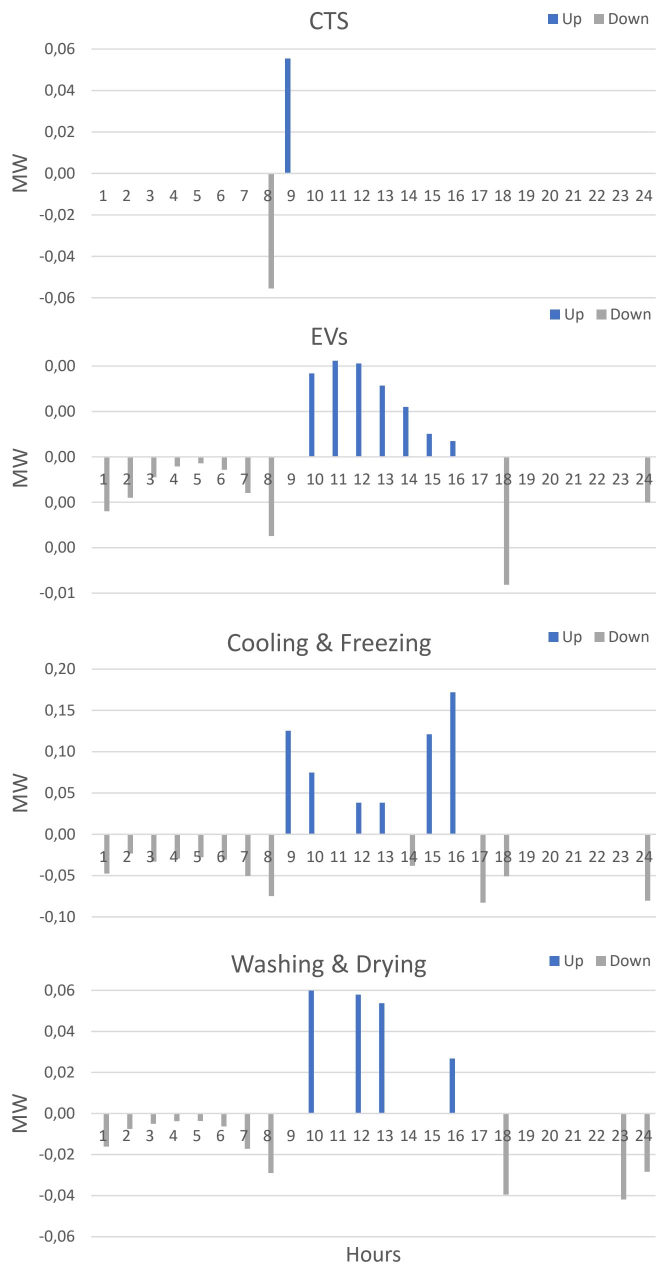


Figure 6: Up and down demand-side management by sector for representative day 4 at bus 6

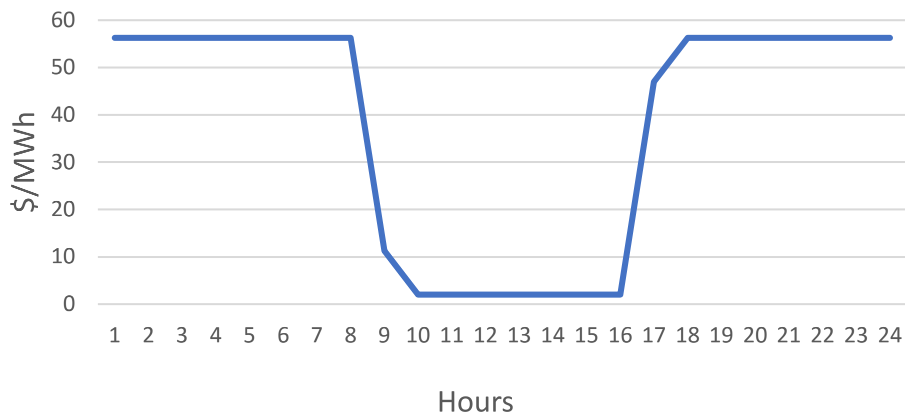


Figure 7: Spot price for representative day 4 at bus 6

DSM patterns, we also look at the spot market price for this day, given in Figure 7, which shows that at midday prices are low and wind is the marginal unit, whereas prices are higher during the remaining hours. Commerce, trade and services (CTS) have a delay time of only one hour and are therefore not used very often. Electric vehicles (EVs) are used more during the day when wind causes low spot prices, and used for down DSM during other hours. Note that in future research we plan to extend the EV DSM formulation to be more specific and to incorporate previously established demand patterns by EV users as we believe this to be more realistic. Cooling and freezing, as well as washing and drying DSM follow a similar pattern.

### *13.3. Hydrogen production*

This version of LEGO also contains the possibility of considering the production of hydrogen via an electrolyzer. We run the rMIP version of LEGO without enforcing renewable penetration and activate the hydrogen option. In this particular case study, we have an annual hydrogen demand of 2190 tonnes that have to be satisfied by one electrolysis plant located at bus 6. We assume that the hydrogen is for industrial use at that bus.

We consider a scenario where energy for the electrolyzer is purchased from the power grid. If we want to determine how clean this hydrogen is, it might be fair to look at the total mix of energy production by technology given in Figure 8, which is 70% CCGTs and 30% clean. The corresponding levelized cost of hydrogen (LCOH) is 2.22 €/kg for this system.

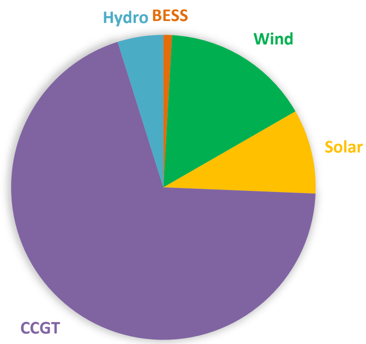


Figure 8: Distribution mix of energy production per technology in power system.

Let us now assume that the industrial demand of hydrogen requires green hydrogen (and not only 30% green hydrogen). In LEGO, we can add a simple constraint limiting the energy consumption of the electrolyzer by the solar production installed in bus 6. This could be interpreted as locally produced, 100% green hydrogen. Running LEGO with this additional constraint yields a LCOH of 2.4 €/kg. The increase in the LCOH here is caused by the reduced[24] overall hydrogen production due to the limited availability of the intermittent solar resource.

In future research we plan to extend the hydrogen sector to be able to analyze all different types of hydrogen, i.e., green, blue or grey hydrogen.

*13.4. Transmission expansion planning DC versus AC*

This section shows the results of a MIP version of LEGO, enforcing a 100% renewable penetration and enabling transmission expansion planning under two

[24]There is a 12% reduction in total annual hydrogen production for the green hydrogen case.

paradigms: a DC- and an AC-OPF setting. The network is shown in Figure 9.

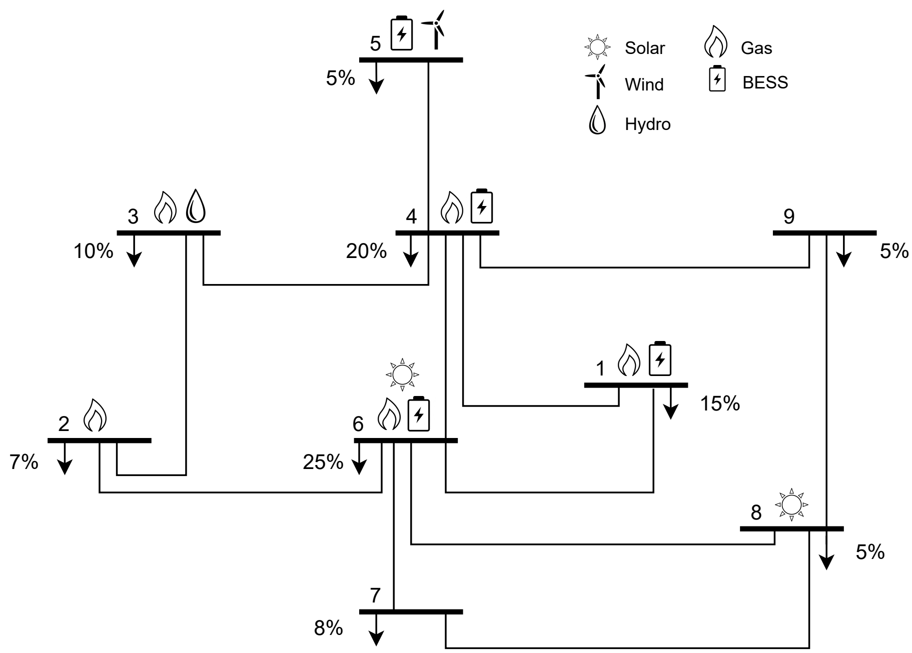


Figure 9: Stylized power system with existing (hydro) and candidate (everything else) generation technologies. Active system demand per bus indicated in percentages.

In this particular data set, the candidate line is built under both DC and AC assumptions (a second circuit between the busses 4 and 5) since building the line allows installing more cheap renewable energy at node 5. And yet, generation investments, power flow, etc., differ significantly. One of the most interesting results is related to the amount of wind capacity installed. Initially, one might be inclined to believe that - since both models build the candidate line between bus 4 and 5 - the newly installed wind capacity at the isolated radial node 5 (where the wind resources are) would also be the same. As shown in Table 56, under the DC-OPF assumption, 7.3 GW of wind capacity is built at bus 5, whereas the AC model only builds 5.8 GW. This result can be explained by analyzing power flow results in more detail. In the DC model, the power flow in line 4-5 is at maximum capacity most of the time. In 3 out of 7 representative days, the line is at maximum capacity for all 24 hours of the day. The flow direction is mostly from bus 5, where the wind is, to the rest of the system via bus 4.

| | **BESS** | **Wind** | **Solar** | **FACTS** |
|---|---|---|---|---|
| DC-OPF (GW) | 7.6 | 7.3 | 12.2 | |
| AC-OPF (GW) | 8.0 | 5.8 | 15.2 | 9 **devices** |
| AC-OPF-tight (GW) | 8.3 | 5.1 | 16.3 | 9 **devices** |

Table 56: Investments (GW) in 100% renewable system under an AC- and a DC-OPF approach.

However, the DC model does not account for voltage limits nor reactive power. Therefore, when doing so with the AC model, power flow results on this line change drastically. For example, active power flow on line 4-5 never exceeds 85% of line capacity because it also has to account for reactive power flow. Moreover, and more importantly, respecting voltage limits becomes essential here. For instance, transporting large amounts of active power from bus 5 to bus 4 causes a voltage drop in bus 5. In order to be within the established voltage limits, which we established between 0.9 and 1.1 p.u., voltages in the remaining system (buses 4 etc.) need to be as high as possible but without exceeding limits. These physical limitations cause that the actual active power flow on line 4-5 is, in reality, closer to 85% of maximum capacity (instead of the 100% predicted by the DC-OPF). As a matter of fact, if we tighten the initially established voltage limits to 0.94 and 1.06 p.u. instead, we observe that even less wind capacity can be evacuated via line 4-5 and the distortions concerning the DC-OPF become even more significant.

The total dispatched active power by technology, presented in Table 57, also reflects the differences between the AC and DC. In DC, 9315 GWh of wind production are assumed when in reality, the AC model shows that only 7548 GWh - so 20% less, are cost optimal due to line limitations. For tighter voltage constraints annual wind production would be around 30% less than what is predicted under the DC model. This aggregated result also reflects the impact on the active power flows. Figure 10 shows the detailed active power production per technology for two representative days (number 7 and number 4).

The AC dispatch also considers the reactive power flows through the network, limiting the maximum active power that it is possible to transfer between nodes as explained in the previous paragraph for the particular case in line 4-5.

| | **BESS** | **Wind** | **Solar** | **Hydro** |
|---|---|---|---|---|
| DC-OPF (GWh) | 8837 | 9315 | 21689 | 1545 |
| AC-OPF (GWh) | 10096 | 7548 | 24121 | 1545 |
| AC-OPF-tight (GWh) | 10561 | 6683 | 24990 | 1545 |

Table 57: Annual production (GWh) in 100% renewable system under an AC- and a DC-OPF approach.

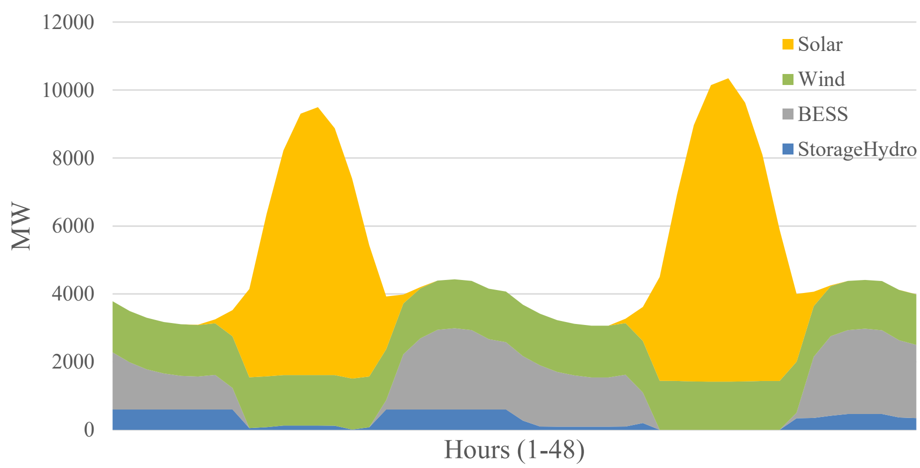


Figure 10: Hourly production (MW) per technology in the 100% renewable system for representative days 7 and 4.

Another difference between the AC and the DC model is the explicit formulation of provision of reactive power. Especially in a 100% renewable system, a-priori, no generator provides reactive power, which the AC-OPF solves by installing 9 FACTs devices. Moreover, the optimal system determined under a DC setting is AC-infeasible due to the lack of provision of reactive power since reactive power is not explicitly accounted for in the DC formulation. When allowing for the ex-post investment of additional FACTS devices, the DC system can be made AC-feasible at an additional system cost of 85 M€. However, such an approach leads to a sub-optimal generation mix.

### References


[1] The Biden plan to build a modern, sustainable infrastructure and an equitable clean energy future, `https://joebiden.com/clean-energy/`.

[2] The European Commission, Communication of the commission. A clean planet for all, The European Commission, 2018.

[3] S. C. Bhattacharyya, G. R. Timilsina, A review of energy system models, International Journal of Energy Sector Management.

[4] General algebraic modeling system (GAMS), https://www.gams.com/.

[5] S. Wogrin, D. Tejada-Arango, R. Gaugl, T. Klatzer, U. Bachhiesl, Lego: The open-source low-carbon expansion generation optimization model (under review)., SoftwareX.

[6] L. Reichenberg, A. S. Siddiqui, S. Wogrin, Policy implications of downscaling the time dimension in power system planning models to represent variability in renewable output, Energy 159 (2018) 870–877.

[7] S. Wogrin, E. Centeno, J. Barquin, Generation capacity expansion in liberalized electricity markets: A stochastic mpec approach, IEEE Transactions on Power Systems 26 (4) (2011) 2526–2532.

[8] S. Wogrin, D. Tejada-Arango, S. Delikaraoglou, A. Botterud, Assessing the impact of inertia and reactive power constraints in generation expansion planning, Applied Energy 280 (2020) 115925.

[9] T. Gerres, J. P. C. Ávila, F. M. Martínez, M. R. Abbad, R. C. Arín, Á. S. Miralles, Rethinking the electricity market design: Remuneration

mechanisms to reach high res shares. results from a spanish case study, Energy Policy 129 (2019) 1320–1330.

[10] S. H. Low, Convex relaxation of optimal power flow—part i: Formulations and equivalence, IEEE Transactions on Control of Network Systems 1 (1) (2014) 15–27.

[11] S. H. Low, Convex relaxation of optimal power flow—part ii: Exactness, IEEE Transactions on Control of Network Systems 1 (2) (2014) 177–189.

[12] J. D. Jenkins, N. A. Sepulveda, Enhanced decision support for a changing electricity landscape: the genx configurable electricity resource capacity expansion model.

[13] D. A. Tejada-Arango, M. Domeshek, S. Wogrin, E. Centeno, Enhanced representative days and system states modeling for energy storage investment analysis, IEEE Transactions on Power Systems 33 (6) (2018) 6534–6544.

[14] S. Gonzato, K. Bruninx, E. Delarue, Long term storage in generation expansion planning models with a reduced temporal scope, Applied Energy 298 (2021) 117168.

[15] L. Hirth, The european electricity market model emma, Neon Neue Energieökonomik GmbH, Tech. Rep. Version (2017) 07–12.

[16] A. Ramos, opentepes mathematical formulation, https://opentepes.readthedocs.io/en/latest/MathematicalFormulation.html.

[17] M. Doquet, R. Gonzalez, S. Lepy, E. Momot, F. Verrier, A new tool for adequacy reporting of electric systems: Antares, CIGRE 2008 session, paper C1-305, Paris.

[18] T. Brown, J. Hörsch, D. Schlachtberger, Pypsa: Python for power system analysis, arXiv preprint arXiv:1707.09913.

[19] J. Johnston, R. Henriquez-Auba, B. Maluenda, M. Fripp, Switch 2.0: A modern platform for planning high-renewable power systems, SoftwareX 10 (2019) 100251.

[20] SpineOpt.
URL https://spine-project.github.io/SpineOpt.jl/latest/

[21] J. Aghaei, M.-I. Alizadeh, Demand response in smart electricity grids equipped with renewable energy sources: A review, Renewable and Sustainable Energy Reviews 18 (2013) 64–72.

[22] G. Strbac, Demand side management: Benefits and challenges, Energy policy 36 (12) (2008) 4419–4426.

[23] L. Göransson, J. Goop, T. Unger, M. Odenberger, F. Johnsson, Linkages between demand-side management and congestion in the european electricity transmission system, Energy 69 (2014) 860–872.

[24] A. Zerrahn, W.-P. Schill, On the representation of demand-side management in power system models, Energy 84 (2015) 840–845.

[25] J. Haas, F. Cebulla, W. Nowak, C. Rahmann, R. Palma-Behnke, A multi-service approach for planning the optimal mix of energy storage technologies in a fully-renewable power supply, Energy Convers. Manag. 178 (2018) 355–368. doi:10.1016/j.enconman.2018.09.087.

[26] X. Zhang, K. W. Chan, H. Wang, J. Hu, B. Zhou, Y. Zhang, J. Qiu, Game-theoretic planning for integrated energy system with independent participants considering ancillary services of power-to-gas stations, Energy 176 (2019) 249–264. doi:10.1016/j.energy.2019.03.154.

[27] Y. Tao, J. Qiu, S. Lai, J. Zhao, Y. Xue, Carbon-Oriented Electricity Network Planning and Transformation, IEEE Trans. Power Syst. 36 (2) (2021) 1034–1048. doi:10.1109/TPWRS.2020.3016668.