# Modelling of 100% Renewable Energy Systems in Integrated Assessment Models by multi-timeframe regression analysis


Parrado-Hernando, Gonzalo [*]
Department of System Engineering and Automation
School of Industrial Engineering, University of Valladolid, Spain
e-mail: gonzalo.parrado@uva.es

Pfeifer, Antun
Department of Energy, Power Engineering and Environment
Faculty of Mechanical Engineering and Naval Architecture
University of Zagreb, Zagreb, Croatia
e-mail: Antun.Pfeifer@fsb.hr

Herc, Luka
Department of Energy, Power Engineering and Environment
Faculty of Mechanical Engineering and Naval Architecture
University of Zagreb, Zagreb, Croatia
e-mail: luka.herc@frodo.fsb.hr

Gjorgievski, Vladimir
Faculty of Electrical Engineering and Information Technologies, University Ss Cyril and Methodius, Skopje, North Macedonia
e-mail: vladimir.gjorgievski@feit.ukim.edu.mk

Batas, Ilija
School of Electrical Engineering
University of Belgrade, Serbia
e-mail: ilija.batas@gmail.com

Duić, Neven
Department of Energy, Power Engineering and Environment
Faculty of Mechanical Engineering and Naval Architecture
University of Zagreb, Zagreb, Croatia
e–mail: Neven.Duic@fsb.hr

Frechoso, Fernando
Department of Electrical Engineering
School of Industrial Engineering, University of Valladolid, Spain
e-mail: frechoso@eii.uva.es

Miguel González, Luis Javier
Department of System Engineering and Automation

[*] Corresponding author.

School of Industrial Engineering, University of Valladolid, Spain
e-mail: ljmiguel@eii.uva.es

Capellán-Perez, Iñigo
Department of System Engineering and Automation
School of Industrial Engineering, University of Valladolid, Spain
e-mail: inigo.capellan@uva.es

## ABSTRACT

Working on holistic approaches that aim to capture a wide range of knowledge, researchers are usually faced with phenomena characterized by different time and geographical scales. This is the case of energy systems and Integrated Assessment Models (IAMs). More specifically, the nature of the variable renewable energy supply (VRES) has traditionally posed a barrier to accurately capturing the effects inflicted by VRES in the energy system.

This research provides a soft link between an energy system model running with an hourly time step, on the one hand, and a yearly-based IAM, on the other hand, by the implementation of an emulator. The proposal here presented is a bridge, based on different types of knowledge, which successfully allows the flow of information between time scales. Results achieve a 100% renewable energy system on a case of Bulgaria. After a brief literature review on the topic, the method is explained in detail, including some results between EnergyPLAN (energy system model) and MEDEAS (Integrated Assessment Model, IAM) for Bulgaria. Results show that the ability of assessment is notably increased from the previous MEDEAS version.

Finally, both results and limitations of this method are discussed. The authors hope this article captures interest in the field of IAMs, especially those which address with energy transition studies.

## KEYWORDS

EnergyPLAN, Integrated Assessment Model (IAM), Energy Transition, Vensim, Variable Renewable Energy Supply.

## INTRODUCTION

Policymaking used to be the principal objective public in the field of integrated assessment modelling. There is a strong relationship between projects and policy measures, as it was recently published in [1]. Nor should this create surprise when energy transition – from a fossil fuel-based economy towards another one supported with large contribution of so-called renewable energies – is finally placed as one of the major challenges of the human development [2][3][4].

In the pathway of energy transition, methods proposing cross-sectoral integration of energy allow high shares of Renewable Energy Sources (RES) in terms of primary energy supply already account a substantial body of research. Most relevant approaches address synergies between power generation, heating, road transport and responsive demand sectors with demand side management. The integration of larger solar-photovoltaic capacities into the Croatian energy system was studied using the synergies between heating and transport sectors [5]. Results showed that higher VRES integrations are easier to be achieved when the system is harmoniously followed up by fostering technologies such as power-to-heat (P2H) and vehicle-to-grid (V2G). In the recent monography, prof. Henrik Lund elaborated renewable heating strategies for reaching a 100% renewable energy solution and grid balancing [6]. Furthermore, distribution of fuels in CHP units can be displaced using different taxing approaches, as shown in [7], through multi-objective optimization, which in turn enables for large power-to-heat implementation.

Energy transition from coal-based towards renewable-based Kosovo's energy system was analysed in [8], with emphasis on P2H technologies. Decarbonisation of an integrated energy system of Italy by 2050 was analysed in [9]. Authors concluded that the whole spread of technologies – cogeneration, trigeneration, V2G, P2H and thermal energy storage – would be required for that transition goal. According to a recent review [10] of best practice examples in P2H implementation, the influence of economic and policy framework factors on the implementation of P2H as demand response emerged as a larger issue compared to the technological development. A number of researchers addressed the flexible operation of the system in the last steps of energy transition, namely the issues of electrification of fuels, producing electro-fuels, synthetic fuels from biofuels and captured $CO_2$ and similar applications [11]. In [12], a zero-emission pathway for the Nordic and Baltic European region was investigated and modelled, concluding that an scenario of a high share of VRES and sector-coupling capacities would be the most economically feasible way forwards. Also, energy system optimizations indicated that most of the investments required for the zero-emission pathway until 2050 would take place already by 2030.

Environmental and economic indicators were used in [13], on the basis of results from the HOMER energy planning tool. Results showed that systems with hybrid storage (electricity and hydrogen) can achieve adequate and reliable decarbonized transport systems while increasing independency in a small island and optimizing the economic and environment sides of sustainability.

The Power-to-Gas concept (P2G) was investigated in [14] to analyse the performance of such innovative storage system. A possibility to integrate co-electrolysers and high temperature methanation was demonstrated, resulting in energy savings. In [15], a decision-making tool for determining the most sustainable use of biomass for carbon management was investigated. The mathematical principles were based on break-even analysis. Emissions-Cost Nexus was considered in identifying the most sustainable pathway using biomass under different baseline conditions. Electrodialysis and hybrid power plants (solar-PV or wind farms) were coupled in [16]. Such hybrid plants are of very attractive flexibility since they also increase the stability of electricity generation. At the same time, electrodialysis was claimed to be a more flexible process compared to reverse osmosis. Results showed that the electrodialysis process was suitable for the integration as a storage within polygeneration systems.

In terms of modelling approaches and scenario analyses, different approaches can be observed. In [17], France was modelled and contrasted under some scenarios, from 0% to 100% renewable share in power production by 2050. Authors tested different configurations of VRES: production, imports, demand flexibility and biomass potential. It was shown that high renewable energy penetration would need significant investments in new capacities, new flexibility options along with imports and demand-response strategies. In the same supply side, it is likely to deteriorate power system reliability whether no technologies dedicated to this issue were installed.

On the economic side, [18] investigated components of the levelized cost of energy (LCOE), emphasizing the idea that external costs of power generation technologies were neglected in the past. Since LCOE was a critical indicator for policy and decision makers, the authors juxtaposed actual costs of renewable and conventional power generation technologies. The same authors internalized some of these external and GHG emission costs across various power generation and storage technologies in all the G20 countries, as they account for 85% of global power consumption. Results showed that renewables were far cheaper than fossil and nuclear sources by 2030, providing statistically display that all the G20 countries had the opportunity to decrease their energy costs significantly. Furthermore, in [19], the marginal prices forecasting method was developed for future energy planning models. The presented “K-SVR” method required also significantly less computational time compared to best known models. In another article, the paradox of energy transition was found in the falling prices of energy. To offer better future electricity prices forecast, the authors proposed a modelling for prices from the residual load obtained by subtracting non-flexible productions from the power load [20].

According to [21], where the IAM “MESSAGE” was used to study the role of hydrogen and storage technologies in a low-carbon energy transition, large VRES shares were supported by the deployment of low-carbon flexible technologies such as hydrogen combustion turbines and concentrating solar power (CSP) with thermal storage. The importance of analysing this kind of flexibility options was also emphasized in [22]. This last study examined an extended an open source energy system model (OSeMOSYS), simulating the operating reserve and related investments for Ireland. That case study covered the effects of linking a long-term energy system model (TIMES) with a unit commitment and dispatch model (PLEXOS). Results showed that investment mismatches decreased from 21.4% to 5.0%. Energy planning processes may be automatized to show deviations in annualize total costs from the optimal energy system structure as [23] stated for an study in the Republic of Serbia.

The country of interest for this document, Bulgaria, was previously modelled to determine what renewable targets would be realistic until 2030 [24]. They used LEAP software and the multi-criteria evaluation method AMS – previously described in [25]. Three scenarios aiming to different RES targets for 2030 supported by different policy mixtures were developed and simulated. Results and related official information were used as inputs. The AMS outputs allowed the identification of the most appropriate scenario for the country. However, this method did not allow hourly analysis.

Regarding IAMs, [26] developed a framework consisting of 18 features to be considered when modelling the dynamics of the power system in order to provide with useful information in high-VRES scenarios, after which a review of novel modelling approaches was done. According to the results, new modelling approaches represented different emerging features but there was a need for further research on inclusion of synergies and for decarbonizing other sectors of the energy system.

Methodologies and methods have being constantly improved, rendering more and more dimensions and feedbacks within complex systems as climate and human-economy metabolism are. Among the tools used to study and assess about such complex systems, the evolution and relevance of integrated assessment models have presented them on the top of policymaking activities [27], not without criticisms [28]. IPCC and global governance and assessment seem to be the natural niche areas of a new branch of science known as "Limits to Growth" or ecology, whose highway started in 70's. Such models contain some advantages regarding holistic – as philosophical meaning – measure to be implemented to improve the whole system, taking into account feedbacks between parts. Global and long-time scales, as well as top-down methodologies are usually common features of IAMs.

Among the methods we found in this short literature review, Residual Load Duration Curve (RLDC) was perhaps the most implemented one in IAMs. The first reference of this method can be read in [29]. Later on, some of the original authors compared their approach with the previous state-of-the-art modelling [26], showing their method had fairly influenced in IAMs (AIM/CGE, IMAGE, MESSAGE, POLES, and REMIND) at different levels of importance to *"describe the fundamental dynamics of the power sector and the effect of VRE"*.

The core idea of this approach was the representation of the influence of variable renewable energies into the shape which was drawn by subtracting – step by step in the data – the production of these supply technologies from the electricity demand (see Figure 1). After that, the curve was sorted into some strategical parts, conforming to supply technologies in the positive axis (underproduction), and power-to-X technologies in the negative quadrant (overproduction). So, given the curve, it was supposed availability of the different technologies at each part, partially losing the hourly management of either over- or underproduction.

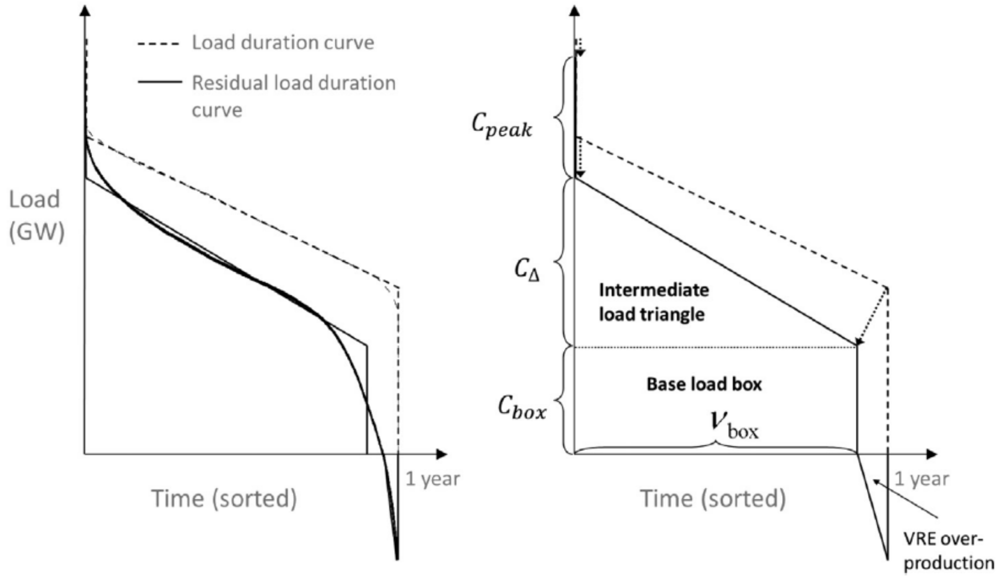



Figure 1. Schematic explanation about the concept of residual load duration curve in the article [29]. The right figure is a simplification of the left one, which is created from substracting the VRES production from the electriticy demand each time step in the data.

The same reference [26] pointed out other methods such as multinomial logit based on LCOE to represent competition between suppliers – IAMs like AIM/CGE and POLES – and exogenous equations with fixed parameters – the case of WITCH.
By looking for global models covering  time horizons +50 years, a good review article [30] saved linear programming, mixed integer programming, partial equilibrium, time slides, computable general equilibrium (AURORAxmp, ETM, GCAM, GEM-E3, MESSAGE, POLES, ReMIND, WEM, and WITCH). Nor did they be related to the system dynamics methodology.

Another kind of strategy was based on soft-linking, i.e., not a direct integration of code but the coupling of two different models to deliver a deeper insight in results than they could by separately. This kind of approaches are currently common between bottom-up models, e.g., between PLEXOS and TIMES [31]. Systems dynamics (SD) software was linked with a load profile model to study the dynamics of the electricity demand in rural areas [32]. However, the literature review we have covered for this article did not include studies of soft-linking between a top-down IAM like MEDEAS – or its next generation, WILIAM – and energy or power flow models.

In order to have a relevant fact in the summary of this introduction, the national climate and energy integrated plan (PNIEC as Spanish acronym) used the bottom-up TIMES-Spain model to develop the policies facing 2030. This model was based on a whole region – Spain – with twelve time slices (four seasons, and three daily periods, peak, night and day) (Section B.1.1. *“Modelo TIMES-Sinergia de la DGPEM”*, [33]).

Two simplification of the complex dynamics produced within the power system were the starting approach in MEDEAS (the IAM of this research). The first simplification regarded the non-VRES side. A polynomial function of order two (see Figure 2) which input was the penetration of variable renewables – wind onshore, wind offshore, solar-photovoltaic (Solar-PV), Concentrated Solar Power (CSP) – into the electricity mix did decrease the capacity factor of dispatchable power plants – hydropower, biogas, bioenergy, geothermal, oceanic, and nuclear power plants. Remaining demand was covered by fossil fuel-based power plants since their capacities were not originally modelled in MEDEAS.

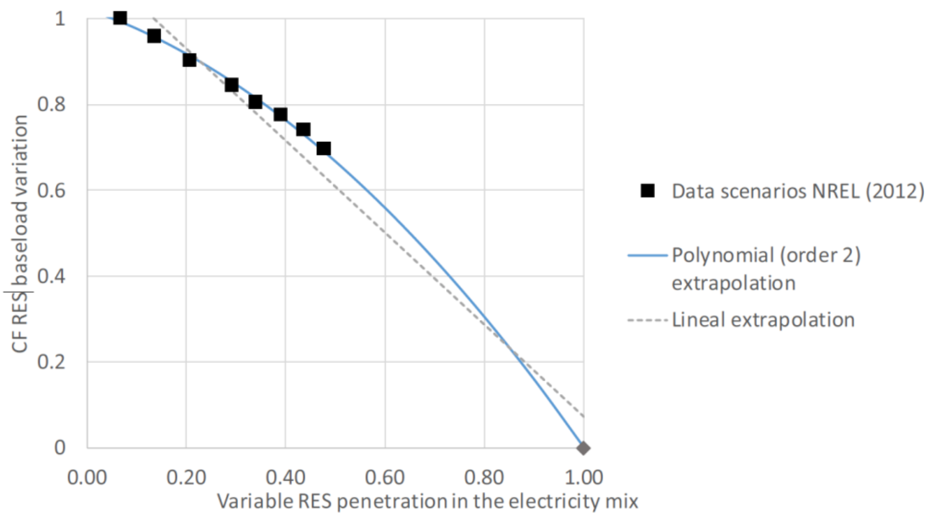



Figure 2. Reduction in the capacity factor of baseload power stations (hydropower, biogas, geothermal, bioenergy, oceanic, and nuclear power plants) due to the penetratino of variable renewable technologies in MEDEAS. This figure was included in the suppplementary material of [34], as figure SM1, created from NREL data (reference 35 of the document). The parabolic equation has the shape $y = -0.6209 * x^2 - 0.3998 * x + 1.0222$ (y is then constained to values between 0 and 1).

On the VRES side, the penetration of these technologies in the electricity mix was translated to estimate their own CF reduction through an exponential function (see Figure 3). So, this second simplification assumed that the whole system deals with curtailment due to overproduction in some hours of the year of the simulation.

Taken into account both simplifications alone, MEDEAS could achieve a total decarbonisation of the power system with an overcapacity of around three times the capacity required to cover the electricity demand without reductions in the CFs, not without sacrificing the other technologies grouped by the MEDEAS concept of 'baseload'. The main "handicap" of this approach was the fact that such baseload facilities (like hydropower plants) experienced higher reductions in the capacity factor than VRES ones, as well as the uncertaintly estimating the overcapacity requirements to achieve 100% power systems.

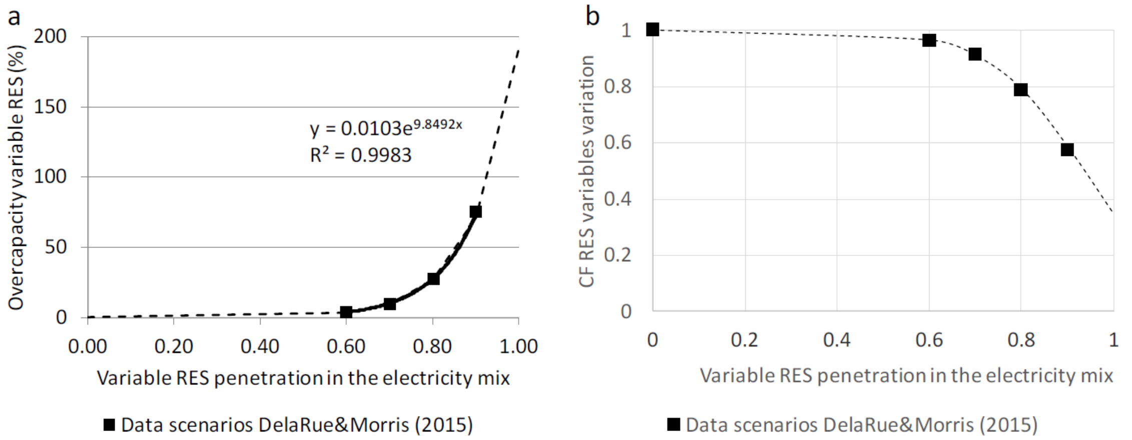



Figure 3. Overcapacity of variable renewable technologies (a, left) and reduction in their capacity factor (b, right) due to the penetration of these technologies in the electricity mix of MEDEAS. The right equation is related to the left one through the equation CF reduction = 1 / (1 + overcapacity). This figure was included in the suppplementary material of [34], as figure SM2, created from the reference 32 of the document.

The work here presented has the principal goal of representing variables of a system which naturally requires a lower time resolution than the unit of a model. Bringing that to our research, the article aims to show the procedure and limitations of an approach to include information from the hourly level to a model running in a yearly basis.
The previously shown method did not involve the effect of flexibility options, which we are going to include in the model and explore further.

## MATERIAL & METHOD

The material of this article may be sorted according to the two model frameworks of Figure 4. On the right hand, the IAM named MEDEAS-Bulgaria requires a wide range of data from economy to energy and demography accounts. The article of reference where MEDEAS is explained is [35]. The Bulgarian version of this model was developed to check the feasibility of the approach carried out on this research.

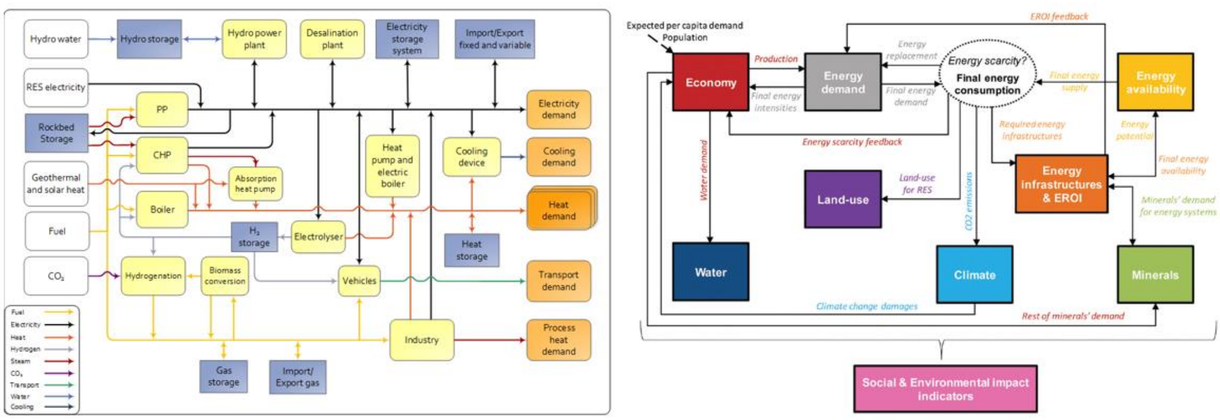



Figure 4. General diagrams of the EnergyPLAN (left) and MEDEAS (right) models, used in this research. The version of EnergyPLAN was 15.1 (15 September 2020), while the version of MEDEAS was MEDEAS_BGR_v12 (Deliverable 4.3).

Hourly resolution was reached by using EnergyPLAN. This energy modelling software was chosen because of its applicability in estimating integration of renewable energy as well as relative simplicity and capability of being automated. Data requirements of EnergyPLAN covers a spread range of the official energy statistics, both in annual – a value – and hourly – a series of values – scales, for both demand and supply sides of the energy system. The documentation of the version used in this research can be read in [36]. Figure 5 summarizes the general flow of steps carried out in the approach. At the right, the subroutines or subtasks define what is done within each step. The first one is the processing of data from different sources to fulfil all the requirements for running both models. Data is analysed to know appropriate technologies delivering flexibility into the regional energy system, as well as the features of the inputs in the next step of permutations (pseudocode in Table 1).

To properly run EnergyPLAN, a technology of interest should not only be defined by its supply side (e.g., capacity and storage) but also by the demand side that concerns to such technology. For instance, the role of power-to-heat (P2H) in EnergyPLAN can briefly be explained. This flexibility option allows for using electricity to produce heat (electric boilers) or to move heat (heat pumps) in two complementary facilities, district grids and the agents grouped as "individual"[†]. Both types of facilities have different profiles of heating and cooling demands. The inputs of P2H are the capacities of boilers (incl. fuel distribution) and heat pumps, the contributions of solar thermal (municipal waste can be used in district grids as well), the storage, and the annual demands of heating and cooling. This complexity is key to understand the main limitation of this approach, which is related to the feasible number of inputs – clusters – when doing the permutations and explained in next paragraphs. Third variables such as conversion factors, efficiencies and capacity factors have influence in the inputs and outputs of the permutations, so they should be also included – as constants or not.

[†] "Individual" refers to suppliers and consumers without district grids connectivity. The agent may encompass households, industries and State sectors.

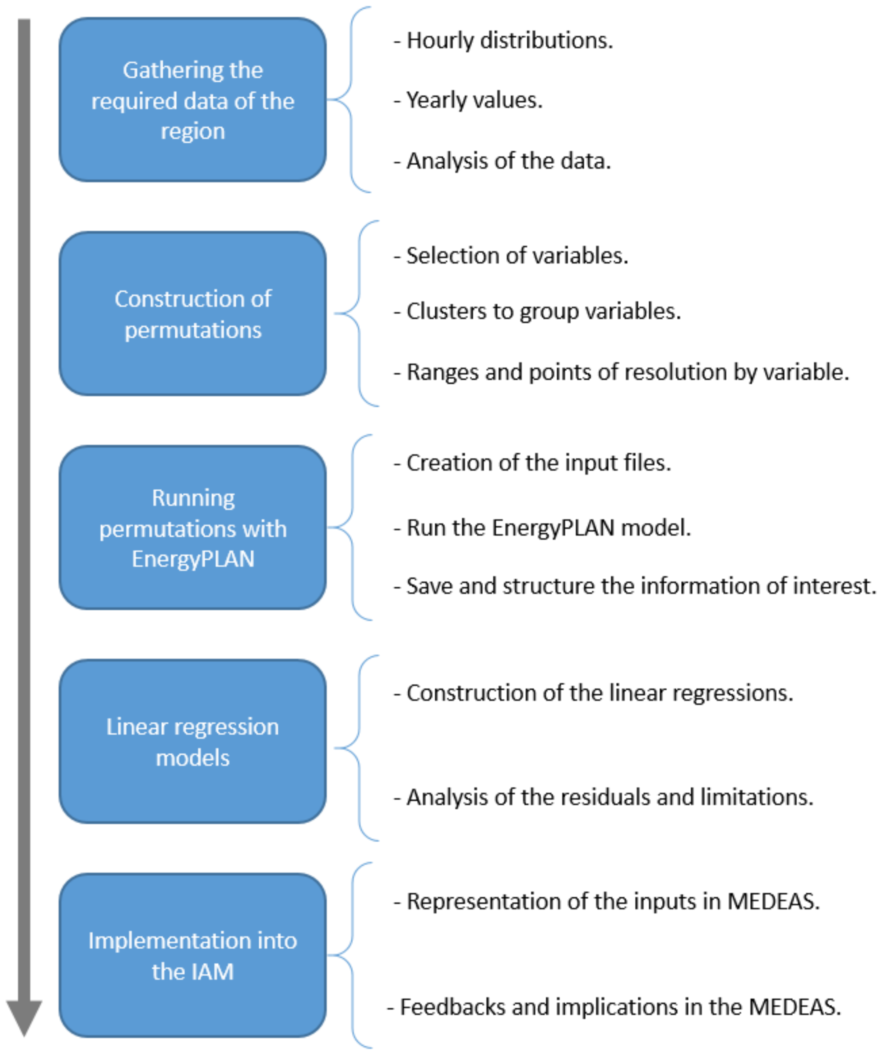



Figure 5. Steps and sub-routines of the approach.

The representation of a basis year for the energy system is necessary to compare results in terms of emissions and share of renewables in final and primary energy consumption, revealing what technologies can do more flexible the system in relation to the intrinsic regional conditions (hourly distributions). Once the base year is ready, authors search for additional data (projections) to determine the values to permutations (inputs of EnergyPLAN). Such permutations are described by means of some features: name, range of values – maximum and minimum –, and number of values in the range (points of resolution).

At this step, a subjective step comes to group inputs into representative clusters. Figure 6 shows the issue why clusters are needed. Without limiting the number of permutations greatly increases the number of permutations and calculation run time. Therefore, clustering serves as a compromise between required run time and output quality. The method here presented would not be feasible in time as of seven variables with ten points of resolution or ten variables with 5 points of resolution. Since we are dealing with subjective steps, the process is iterative up to permutations are decided. The clustering criteria is twofold. A cluster must allow the effect of the represented technology, in such a way that proportionality is assumed within the cluster. Additionally, some technologies may have a similar effect, so all of them may or not be grouped in the same cluster, depending on what is being studied by the modellers. For instance, pump hydropower storage and electric batteries in the power grid both increase the flexibility option so-called "storage" while electric batteries of vehicles rely so different on the demand of road transport that they are considered as another cluster. Some general ideas might be considered to create the clusters:

- Variables allowing the use of a technology that provides with flexibility to the energy system. These variables can be sorted into first, second and third spheres of influence.
- Projections may help to determine what variables should have wide range (according with the units) to reveal the influence of flexibility and what others may be considered as constant.

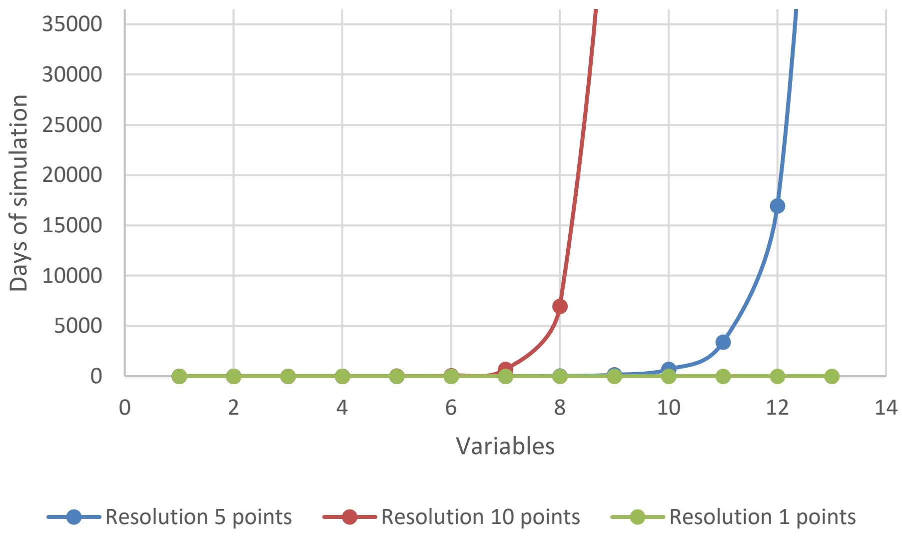



Figure 6. Computational cost in days of the permutation process with EnergyPLAN, according with a number of points by variable and the number of variables. The time of one simulation is assumed to be one second.

We would like to highlight that clusters have specific ranges to do the permutations, so the results are correct as long as the model does work within these ranges.

An example of features structure for the inputs of permutations can be shown in Figure 9, as well as the related clusters in Figure 10. Three points were included to render the input "Electricity smart" (units in TWh), i.e., the electricity demand of electric vehicles able to run in smart mode[‡]. The maximum value corresponds to 100% of the electricity demand running as smart mode. In this case, related parameters such as charging capacity,

[‡] EnergyPLAN has two different modes to represent the hourly simulation of electric vehicles – dump and smart modes. Smart mode allows charging and discharging electricity from and to the power grid to flexibilize the demand side, always satisfying the hourly demand. Further information in the documentation of EnergyPLAN.

battery storage and the use of other fuels were proportionally modified in accordance to the values of the input ("Electricity smart").

Clustering and pre-processing of data is done with MS Excel, while Power Query software facilitated the creation of permutations as well as the post-processing tasks with results (EnergyPLAN's files "output.csv"). Then, a Python code is used to run EnergyPLAN as times as permutations were created. This code translates the values of permutations into input files of EnergyPLAN (code in Table SM 2). Once these input files are generated, EnergyPLAN is run and generated outputs are saved and allocated (code in Table SM 3). EnergyPLAN's warnings[§] are also saved to decide whether or not repeat the permutations.

Table 1. Pseudocode summarizing the process from gathering data up to the creation of the linear regression models.

```
Define input parameters
Define path of the folder containing the input files
Define path where outputs are saved
Open the list of files to be run by EnergyPLAN
For each file in the list:
        Open case in EnergyPLAN
        Save the outputs of interest from the results file
        Clear workspace
Check errors in simulations
If error is detected:
        Run the specific case with a bigger time step between parts of the process.
```

The results of the permutations, i.e., the values of the clusters (inputs) and the EnergyPLAN's results, are used to build up the linear regression models which will represent some hourly features of the system inside the yearly IAM. We create as regressions as outputs we would like to represent. Each of the linear regression are built in the same iterative way. The procedure can be quickly followed in Table 2.

Firstly, the correlation matrix between inputs and the output to be estimated is calculated. The highest score (between 0 and 1) means the variable it refers to is the most correlated among all of them. It is selected to the first linear regression. The raw residuals – difference between the actual and the estimated value– of this first regression model yet saves information about the output, so we did again the correlation matrix however now with the residuals as output. The first input is not correlated with the residuals (score equal to zero), so we select the next highest value. This process is repeated up to all the correlation scores are below a criteria. Then, the final linear regression is made up by superposition, i.e., all the independent terms are added, and the dependent terms are related to its corresponding input variable.

[§] EnergyPLAN might deliver some warnings after the simulation is run: critical excess of electricity production (CEEP), grid stability problem, power plant or import problem, synthetic or biogas shortage, V2G connection too small, and negative electricity demand. Further information in the documentation of EnergyPLAN.

Table 2. Pseudocode explaining the process of building a linear regression model to estimate an output based on permutations.

```
Import values of the permutations –inputs and outputs.
Define the value of correlation score (p) by which the process of creating the regression
model is stopped.
Initialize variables required to the loop
Calculate the correlation matrix with the original output
Select the variable of the highest score
While score > p:
        Create the linear regression model
        Save the independent value of the regression
        Save the dependent values of the regression
        Calculate and save the residuals of the regression
        Calculate the correlation matrix with the residuals as output
        Select the highest score
end of the while loop
Addition of the independent values
Save the information of interest into a file
```

Finally, the meaning of inputs in the IAM should be as similar as possible to the energy model's input. Consistency should be present in this last step of the approach.

## BULGARIA AS CASE OF STUDY. RESULTS & DISCUSSION

Bulgaria was selected as case of study because of both available models, EnergyPLAN (energy model) and MEDEAS (Integrated Assessment Model). The availability of data was another relevant factor to choose for this country. Eurostat's database [37], IEA database [38] and the Query tool (VBA macros) of IRENA [39] were the data sources of this study. Atlas of wind speed [40] and solar irradiation [41] facilitated hourly data to create the distributions of both related renewable technologies.

In order to capture the behaviour of the Bulgarian energy system, the input files of EnergyPLAN were switched 34992 times (permutations). The features of representative clusters were defined in Table 3 and Table SM 1 of the supplementary material.

The selection of clusters relied on the characteristics of the region analysed. Criteria was subjective and based on the expertise of the modellers and assesors. Authors searched through published data and accordingly determined the structure of permutations. Some key sentences were written following this lines to understand the selection carried out for Bulgaria.

- The proyections of run-of-river hydropower capacity does not seem to be relevant. This technology was represented as constant due to its relatively small size and potential in the country.
- Dounabe river –the largest river of Europe – runs along the Bulgarian-Romania border. The gradient of this river is not stimulating for building dam hydropower

plants since the construction would require massive lakes and agreements between both countries. One should take into account that such agreements would be firmed by different national enviromental regulations.

- The distribution of solar irradiation for Bulgaria (Figure 7) provided a potential of up to 100 GW of solar-PV capacity with an average solar iradiation of 1300 $kWh/m^2$. Other technical conditions such as crop yields or protected areas to install solar pannels are out of this research.
- The distribution of wind speed for Bulgaria (Figure 8) provided a great potential of wind power technology in the mountainous regions and the Noth-East of the country, on the shore of the Black sea. Other technical conditions such as maximum slope or sea depth to install wind turbines are out of this research.
- District heating demand represented 12,55 TWh –60 % of total heat demand in Bulgaria, the rest was considered as "individual", supplied by boilers fueled with biomass.

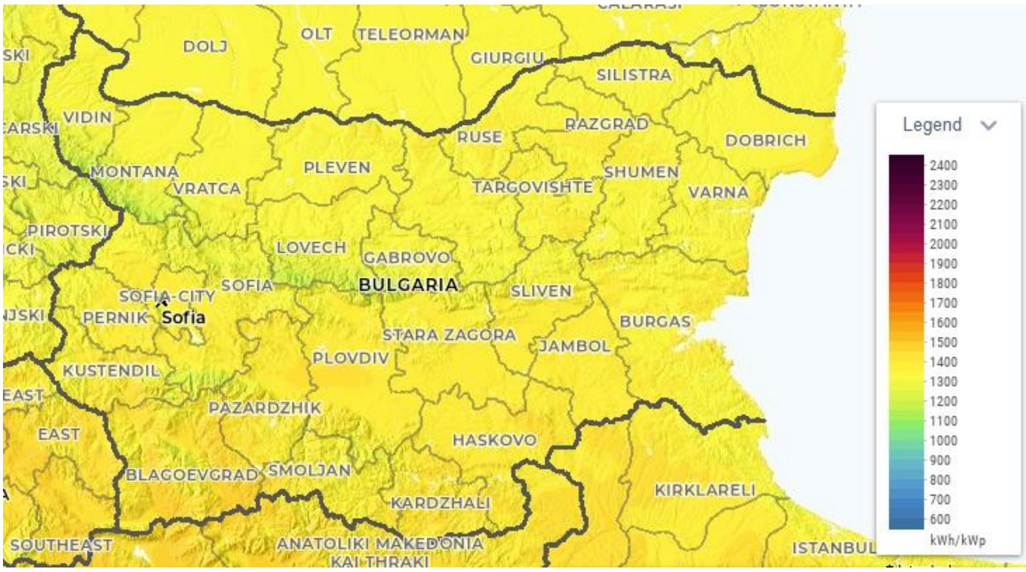



Figure 7. Distribution of solar iradiation for Bulgaria. Source: [42].

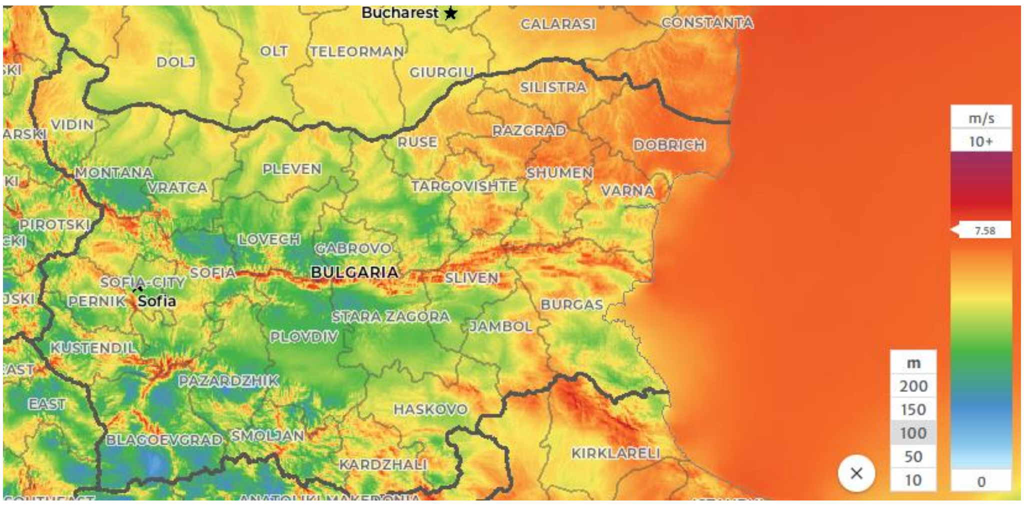



Figure 8. Distribution of average wind speeds for Bulgaria. Source: [40].

Table 3. Definition of the clusters made to do the permutations with EnergyPLAN. All the variables of the same cluster are modified together when the permutations are being generated. More information in the Supplementary Materials.

| | Name of the cluster | Meaning of the cluster | Number of EnergyPLAN's input to represent the cluster | Resolution of the cluster |
|---|---|---|---|---|
| INPUTS | Wind | Wind power plants. | 1 | 3 |
| | SolarPV | Solar-photovoltaic power plants. | 1 | 4 |
| | DamHydro | Dammed hydropower plants. | 1 | 2 |
| | Backing | Power plants with rankine cycle with back-up features. | 2 | 2 |
| | ElecTransport | Electrification of the transport sector. | 13 | 2 |
| | P2H | Power-to-heat. | 2 | 3 |
| | Storage | Storage in the power grid, pump hydropower plants and Rockbed (high temperature) storages. | 6 | 3 |
| | FlexibleDemand | Flexible electricity demand. | 7 | 3 |
| | FossilIndustry | Level of decarbonization in Industry. | 4 | 3 |
| | SynthGas | Use of hydrogen to generate synthetic gas. | 1 | 3 |

Further clarification of clusters, parameters, value of the parameters and their relations is displayed in Figure 9. And Figure 10. On this example, cluster displaying transport electrification is displayed. As can be observed, change of one parameters value influences the change in the rest of the values.

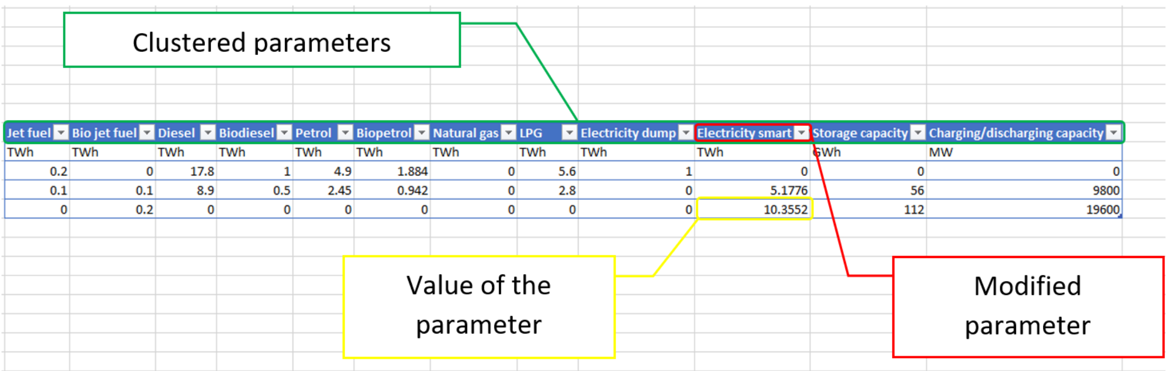



Figure 9. Partial view showing the relations in the features of some inputs for the Bulgarian case of study. An input is given by the name, a range, and the number of points within the range.

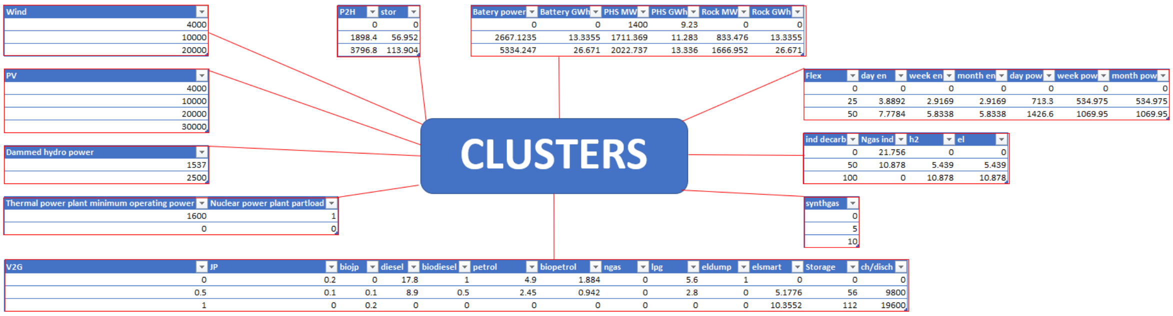



Figure 10. Final structure of inputs and clusters for the Bulgarian permutations. Ten clusters with different resolution of inputs.

As mentioned before, we needed a criteria to stop the procedure when building up the linear regression model. The value selected for our case of study was 0.05, so we selected and create linear regressions up to the correlation scores are below this value.
The outputs of interest with their related linear models are enumerated following this paragraph. The equation to calculate the variation in the capacity factor (input for the regression models) follows the structure *Variation CF = (Maximum CF – Calculated CF) / (Maximum CF)*. The inputs appears in the same order than they are relevant according with the correlation scores.

- Variation in the capacity factor of wind power plants.

$$VarCFwind = 0.741214 + 0.000799 * SolarPV + 0.000957 * Wind + 0.006613 * Baseload - 0.120798 * FossilIndustry - 0.001173 * Storage - 0.749565 * SynthGas - 0.04446 * V2Gstorage$$

- Variation in the capacity factor of solar-photovoltaic power plants.

$$VarCFsolarPV = -0.387804 + 0.000602 * SolarPV + 0.00072 * Wind - 0.105805 * FossilIndustry - 0.745347 * SynthGas + 0.002903 * Baseload - 0.038897 * V2Gstorage - 0.000416 * Storage$$

- Variation in the capacity factor of run-of-river hydropower plants.

$$VarCFRoR = 0.693932 + 0.002626 * Baseload - 0.000614 * Storage + 0.000068 * SolarPV + 0.00101 * FlexibleDemand + 0.009445 * V2Gstorage + 0.08698 * SynthGas - 0.005322 * FossilIndustry$$

- Variation in the capacity factor of dam Hydropower plants.

$$VarCFDamHydro = 0.169749 + 0.000404 * DamHydro - 1.21024678190544E - 07 * SolarPV$$

- Variation in the capacity factor of backing facilities (traditional Rankine-cycle power plants).

$$VarCFPP = 75.508738 - 0.0313042 * Baseload + 0.000476 * SolarPV - 0.11511 * FossilIndustry + 0.000582 * Wind + 0.003385 * FlexibleDemand - 0.343115 * SynthGas - 0.019282 * V2Gstorage$$

- Variation in the capacity factor of combined heat and power plants.

$$VarCFCHP = 26.250991 + 0.016443 * Baseload + 0.001694 * Wind + 0.000901 * SolarPV - 0.139477 * FossilIndustry - 0.000836 * Storage + 0.002362 * FlexibleDemand - 0.301596 * SynthGas$$

- Variation in the capacity factor of nuclear power plants.

$$VarCFCHP = 0.005535 - 5.1878E - 06 * Baseload + 6.0674E - 07 * Wind + 3.7992E - 07 * SolarPV - 0.000090 * FossilIndustry - 0.00068 * SynthGas - 0.000041 * V2Gstorage$$

- Heat losses in storage (TWh).

$$HeatLossesStorage = -3.242717 + 0.000486 * Wind + 0.000313 * SolarPV - 0.003575 * Baseload + 0.000810 * Storage - 0.051809 * FossilIndustry - 0.357013 * SynthGas - 0.021264 * V2Gstorage$$

- Losses of electricity in the storage of electric vehicles (TWh).

$$LossesV2G = -0.256752 + 0.015472 * V2Gstorage + 0.00002 * SolarPV + 0.003939 * FossilIndustry - 0.000162 * Baseload - 0.000205 * FlexibleDemand + 0.024668 * SynthGas$$

At this point, we would like to write some notes about the accuracy of the proposed linear regression models to estimate variables of interest. The variation in the capacity factor of wind and solar-PV power stations was selected to illustrate this discussion. A regression model is a tool to estimate values from evidence materialized on data. The values of the permutations are the evidence to check the level of accuracy for the regressions. When raw residuals are zero, it means the estimated value is the same than the one calculated in EnergyPLAN. Of course, since we are doing clusters of some inputs which may have non-linear relationships, residuals necessarily exists and they are going to be analysed. The difference between the estimated and the permutation values (residuals) are represented in relative probability of occurrence in Figure 11 to show a general view of the accuracy in the regression model. On the right of the same figure, the normal probability plot are represented to show the degree of similarity the residuals seems to be regarding with a normal probability distribution. Looking the right side of Figure 11, we can conclude that normal distributions does not fit well the residuals.

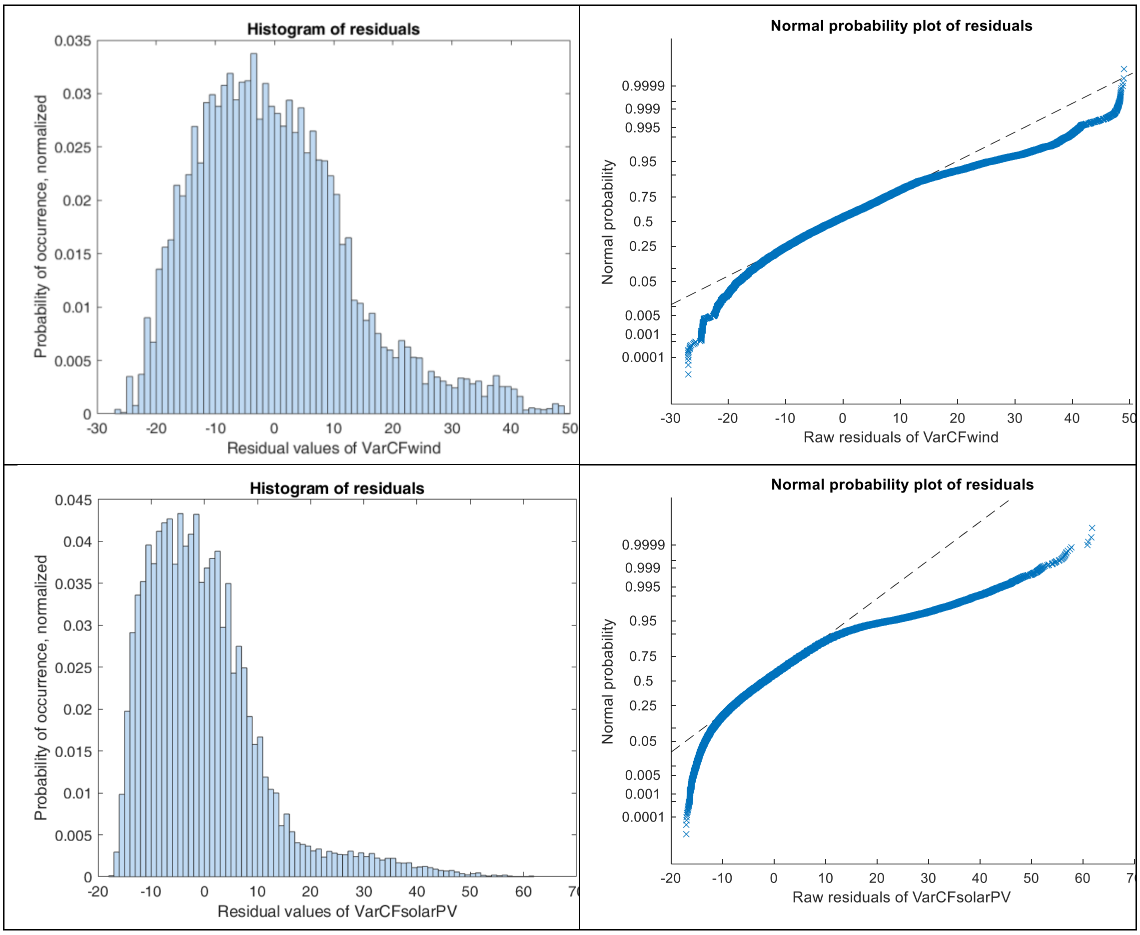


Figure 11. Probability of occurence (relative to the number of permutations) on the left side and normal probability plot on the right, for two outputs: the variation in the capacity factor of solar-photovoltaic –VarCFsolarPV – and wind –VarCFwind – technologies.

Difference between the values of the permutations and the estimated values emerged from regression models of wind and solar-PV technologies is rendered in Figure 12. There were some points for the discussion. Firstly, the linear regression model underestimated the impact of the variability before roughly 60%, provinding less impact on the technology than EnergyPLAN, specialy relevant for solar-PV cases. The maximum absolute difference reached 49% for the case of wind as well as 84% for the case of solar-PV. However, this changes above 60%, when regressions seemed to overestimate such impact into the technology. Secondly, the regression models must be upper constrained by the maximum capacity when implementing the approach in the IAM, since we cannot provide negative variations –it would mean a greater capacity factor than the maximum of that technology. Thirdly, it was surprising the convergence to zero variation (less impact) in both wind and solar figures when increasing the penetration of renewables.

The convergence of capacity factors happened due to the constrains implemented by the authors to model Bulgaria as a closed energy system (without international power connections). When a region is modeled in such a way, a number of simulations may not satisfy system stability requiring emergency import of electricity. In this case, the results with emergency import present are not being considered. Thus, only the cases with succesful integration of RES are being analyzed which means that capacity factor increases with the increase of the share of RES.

Actually, a reduction of more than a percentage in the capacity factor of any technology would lead for reducing investments since profit would be reduced. This is subjected to discussion because of the dynamics of fixed and variable costs of technologies –e.g., LCOE (Levelized Cost Of Energy) of solar-PV is decreasing more and more, from currently 0.142 $/kWh to 0.093-0.128 $/kWh or even less in 2030 [43]. For this study, it was decided to be conservative so a maximum percentage of decline in capacity factors of all variable renewables equal to 20% was included, for which completelly stoped new required capacity. VRES development continued to support the pathway to a decarbonized energy system whether flexibility technologies were fostered.

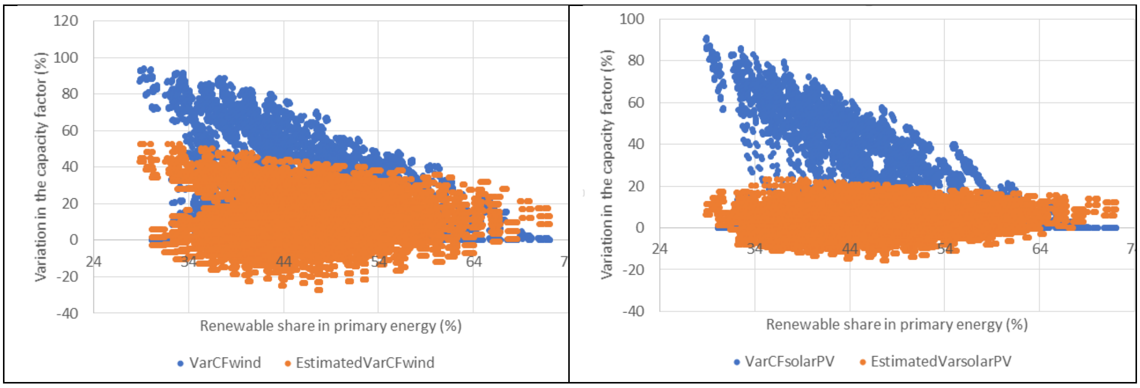


Figure 12. Permutation values and corresponding estimated values for variations in the capacity factors of wind (left) and solar-PV(right) technologies.

Going deeper into detail, greater values corresponded with odd scenarios. For instance, the case of the maximum difference in solar-PV stated a capacity ratio wind/solar-PV of 1/3, the lower value of storage, and zeros in the last three variables (FlexibleDemand, FossilIndustry, and SynthGas). This point was located at the left on the figure.

Almost all inputs of the regression models were represented in MEDEAS-BGR. However, a consistent approach for some technologies was required in order to be coherent with the regular results of this IAM, respecting the general framework of the system.

Synthetic gas was required to model the regressions and firstly included in MEDEAS models. This fuel has been endogenously modelled taking into account the whole chain (25% of global efficiency), from the energy demanded by the economy, checking for available energy resources, and feedback to the economy.

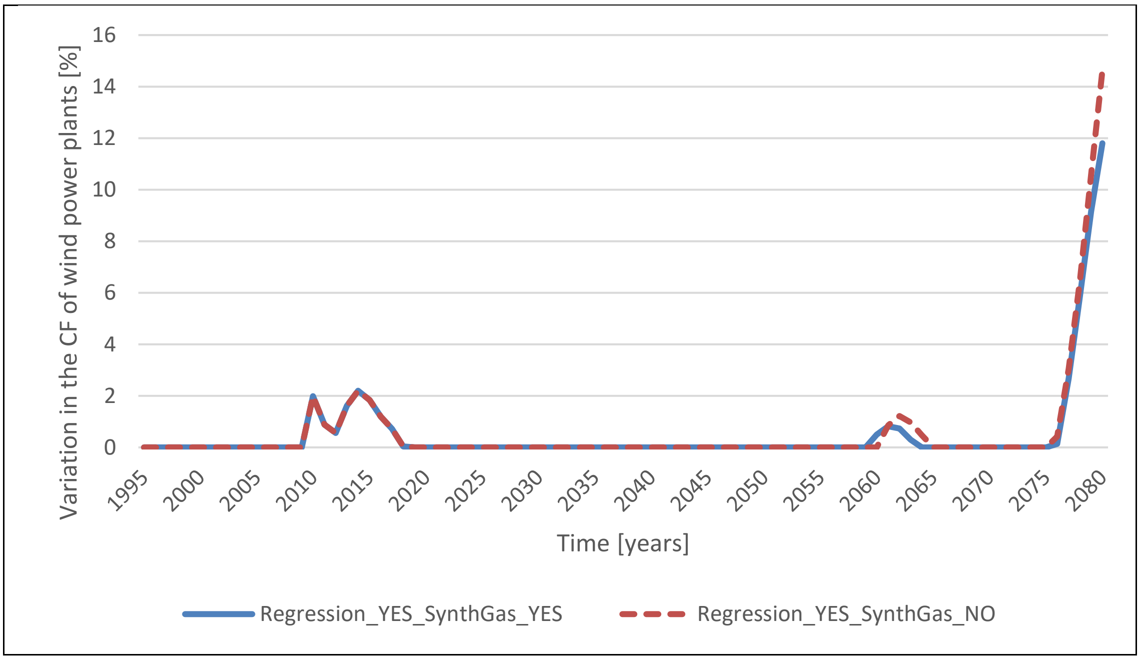
MEDEAS-BGR has five final fuels –electricity, heat, liquids, gases, solids. The demand of this hydrogen-based gas was calculated from the final energy demand of three sectors for gases: “Coke, Refined Petroleum and Nuclear Fuel”, “Chemicals and Chemical

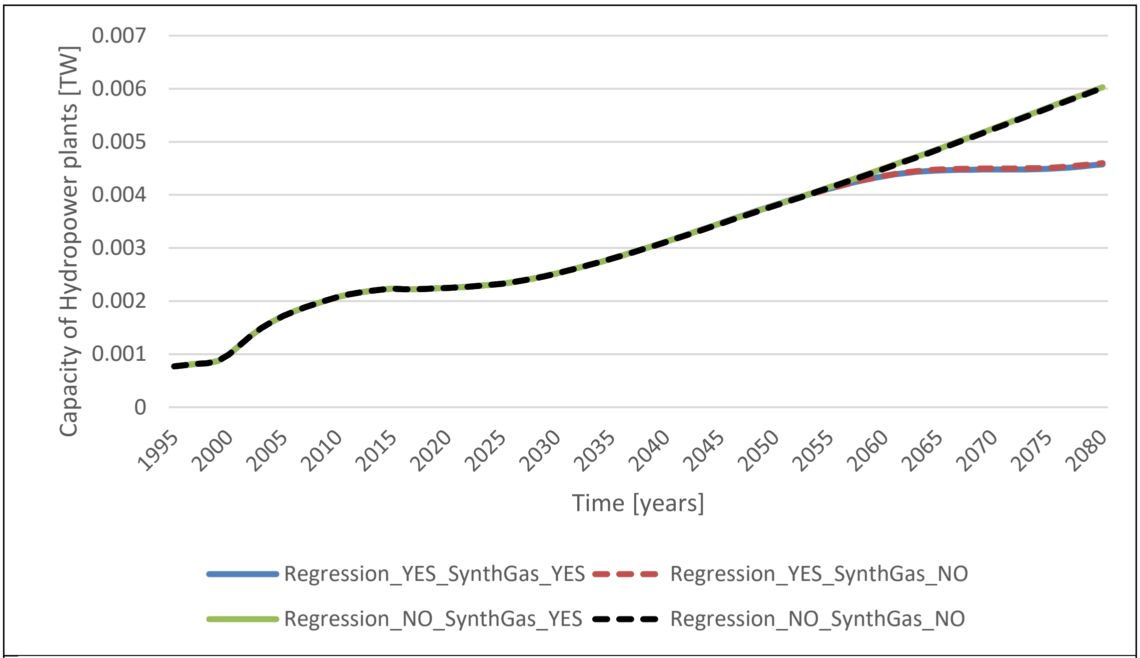
Products”, and “Basic Metals and Fabricated Metal”. So, it was assumed that gases demand of these industries can be replaced by synthetic gas. An exogenous policy progressively increased both the share of substitution and the capacity of Power-to-Gas facilities. The electricity required to produce synthetic gas was added to the total electricity demand.

Since extraction of fossil fuels in MEDEAS-BGR was the last step to fulfil the demand – after biogas and X-to-gas transformations) – natural gas by synthetic gas could not be directly substituted. To solve this issue, synthetic gas was selected as first-priority fuel to satisfy the demand of gases (final energy), which diminished the extraction of natural gas so resulting in a higher security facing possible scarcity of this fossil fuel. The last influence took into account the reduction of GHG emissions. A bottom can either activate or deactivate the consequences of this flexibility option.

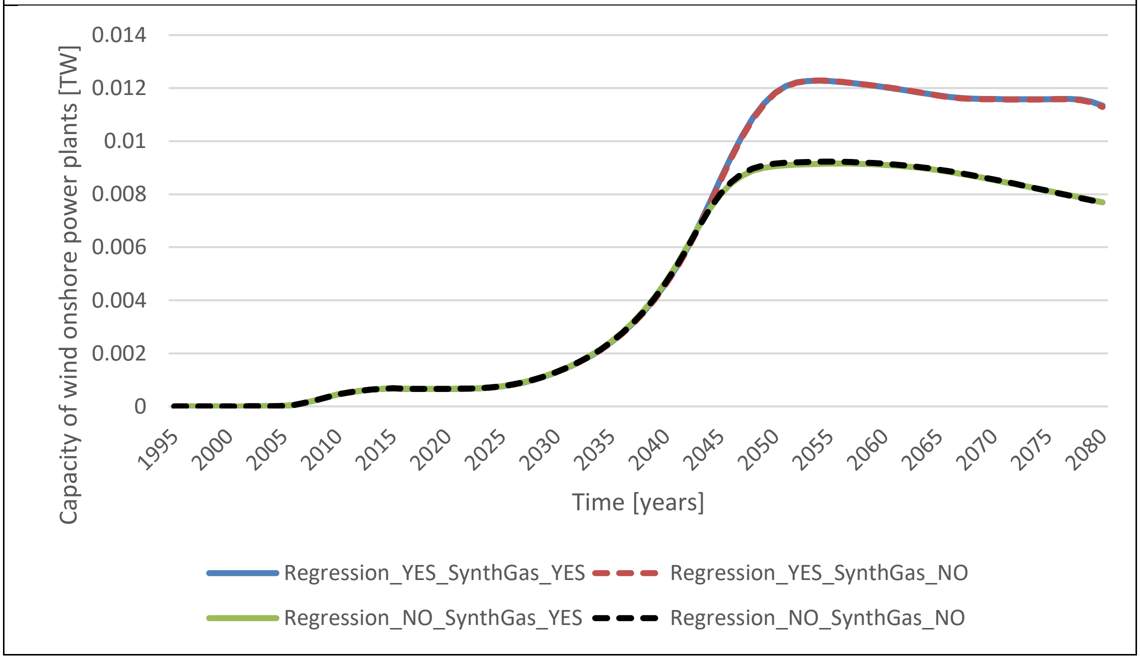
Capacities of thermal power plants (steam power turbines, combined cycle gas turbines, internal engines) were also firstly modelled in Bulgaria. The rest of inputs of the regression models were linked according with variables of same meaning.

The same scenario (Table SM 5) was introduced to show differences between simulations in the Bulgarian version of MEDEAS. The four simulations come from switch on/off two features of the model: method (original and regressions) and synthetic gas (yes or no). Three results of interest are plotted in Figure 14.





0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Capacity of Hydropower plants [TW]
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
Time [years]
Regression_YES_SynthGas_YES
Regression_YES_SynthGas_NO
Regression_NO_SynthGas_YES
Regression_NO_SynthGas_NO

0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
Capacity of wind onshore power plants [TW]
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
Time [years]
Regression_YES_SynthGas_YES
Regression_YES_SynthGas_NO
Regression_NO_SynthGas_YES
Regression_NO_SynthGas_NO

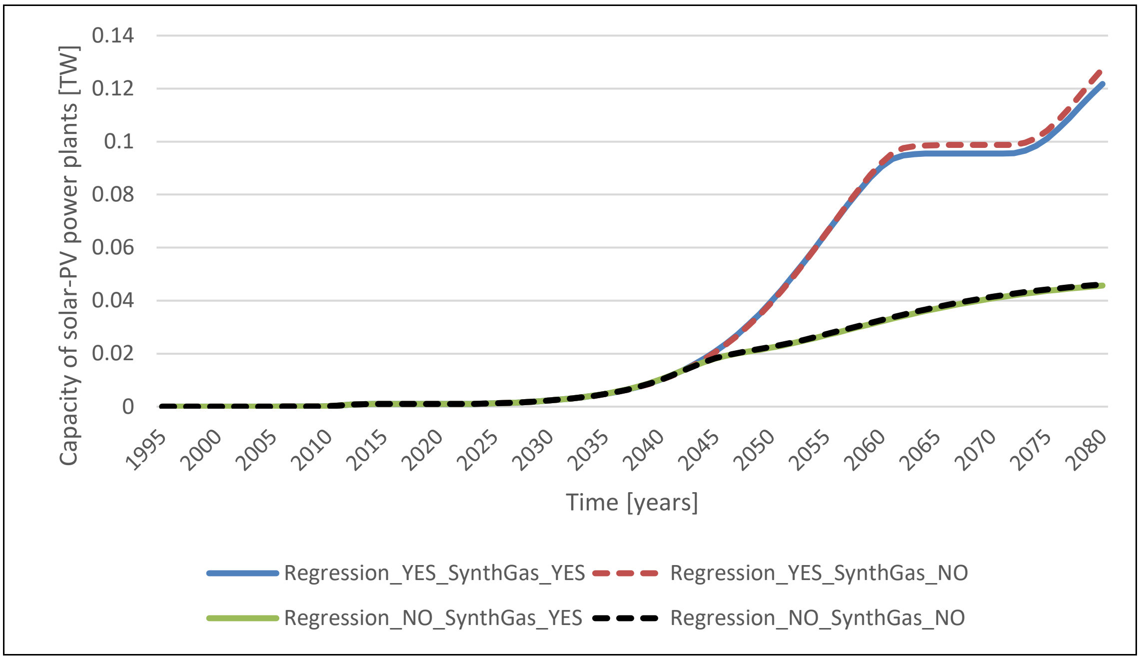
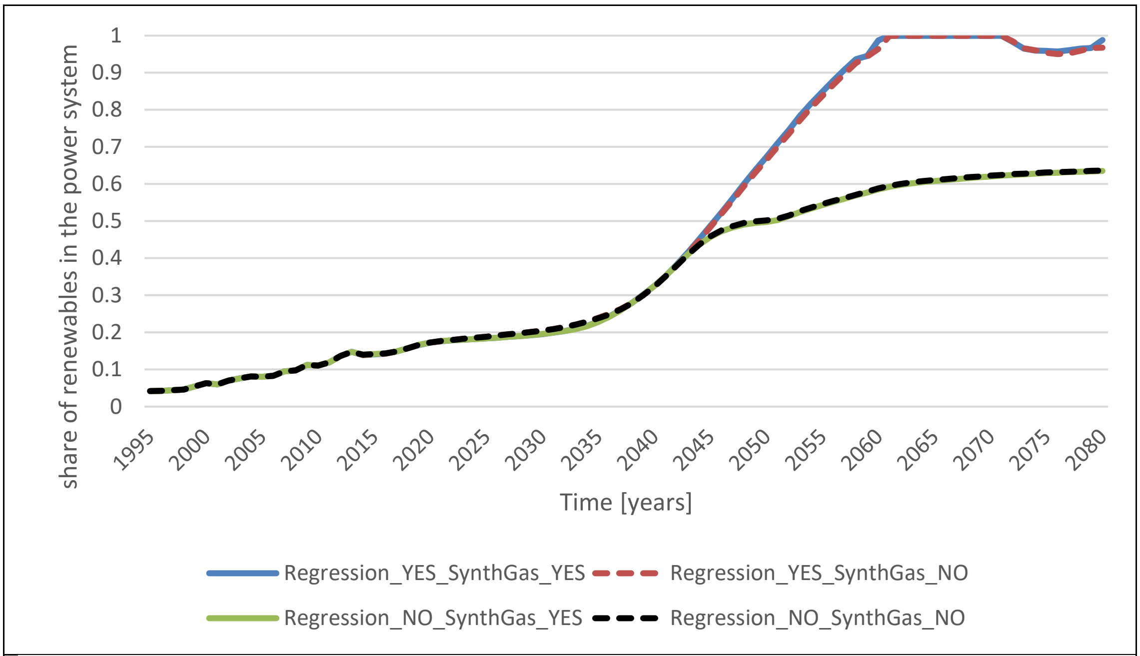
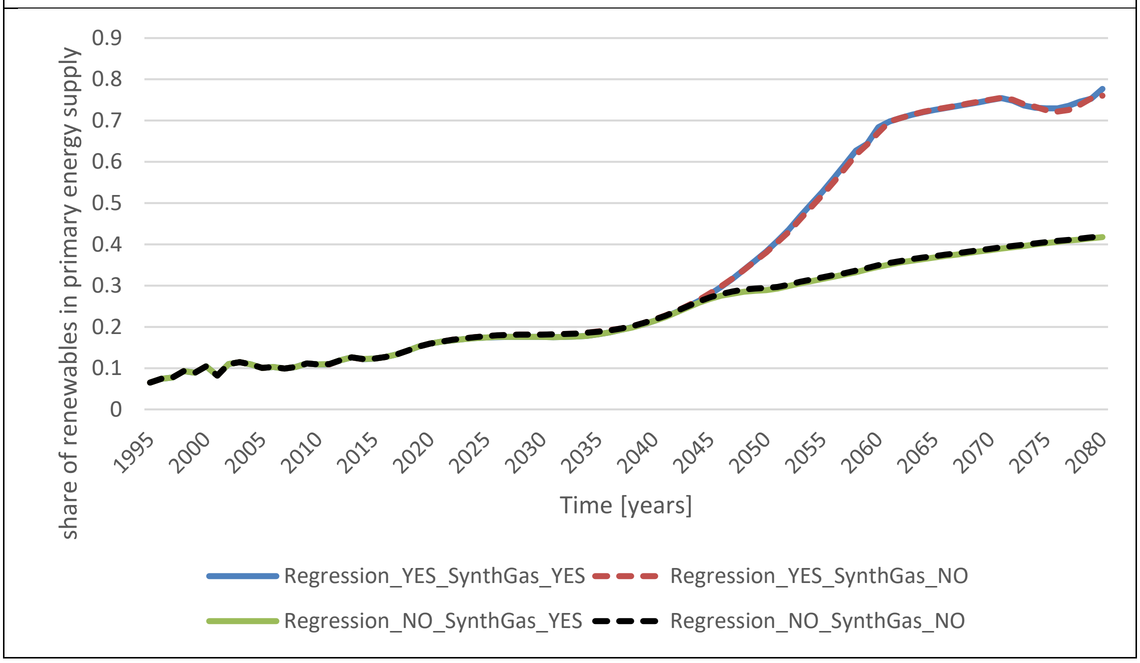


Figure 13. Results in MEDEAS-BGR for the variation in the capacity factor of wind power plants (first), capacity of hydropower plants (second), capacity of wind power plants (third), and capacity of solar-photovoltaic power plants (fourth). Four cases of the same scenario were represented: original method without synthetic gas (black dotted line), original method with synthetic gas (green continuous line), regression method without synthetic gas (red dotted line), and regression method with synthetic gas (blue continuous line).

Analysing the new method (regressions) in the scenario (Figure 13), variation of capacity factor for solar-PV and nuclear units was zero over the simulation. However, technology moved by wind experimented three periods of time with reduction due variability, from 2009 to 2019, from 2060 to 2065, and from 2075 forward, when variation sharply increased. Synthetic gas reduces the impact on this indicator in comparison with the simulation without production of this $H_2$-fuel. As the rest of figures show, the new method allow for more integration of capacities of variable renewables (roughly +29% for wind and +75% for solar-PV in 2050) when comparing with the old version of MEDEAS-BGR, what partially stopped the hydropower deployment since it had less priority to deliver electricity than VRES. Synthetic gas restricted the deployment of VRES a little, due to differences in the EROI of the system. Production of this fuel increased the electricity demand in a poor efficiency process of 25%, which get worse the whole whole energy system to satisfy goods such as pannels and wind turbines. This effect can be shown in the solar-PV figure, which development achieves higher levels than wind.


share of renewables in the power system
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
Time [years]
Regression_YES_SynthGas_YES
Regression_YES_SynthGas_NO
Regression_NO_SynthGas_YES
Regression_NO_SynthGas_NO


share of renewables in primary energy supply
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
Time [years]
Regression_YES_SynthGas_YES
Regression_YES_SynthGas_NO
Regression_NO_SynthGas_YES
Regression_NO_SynthGas_NO

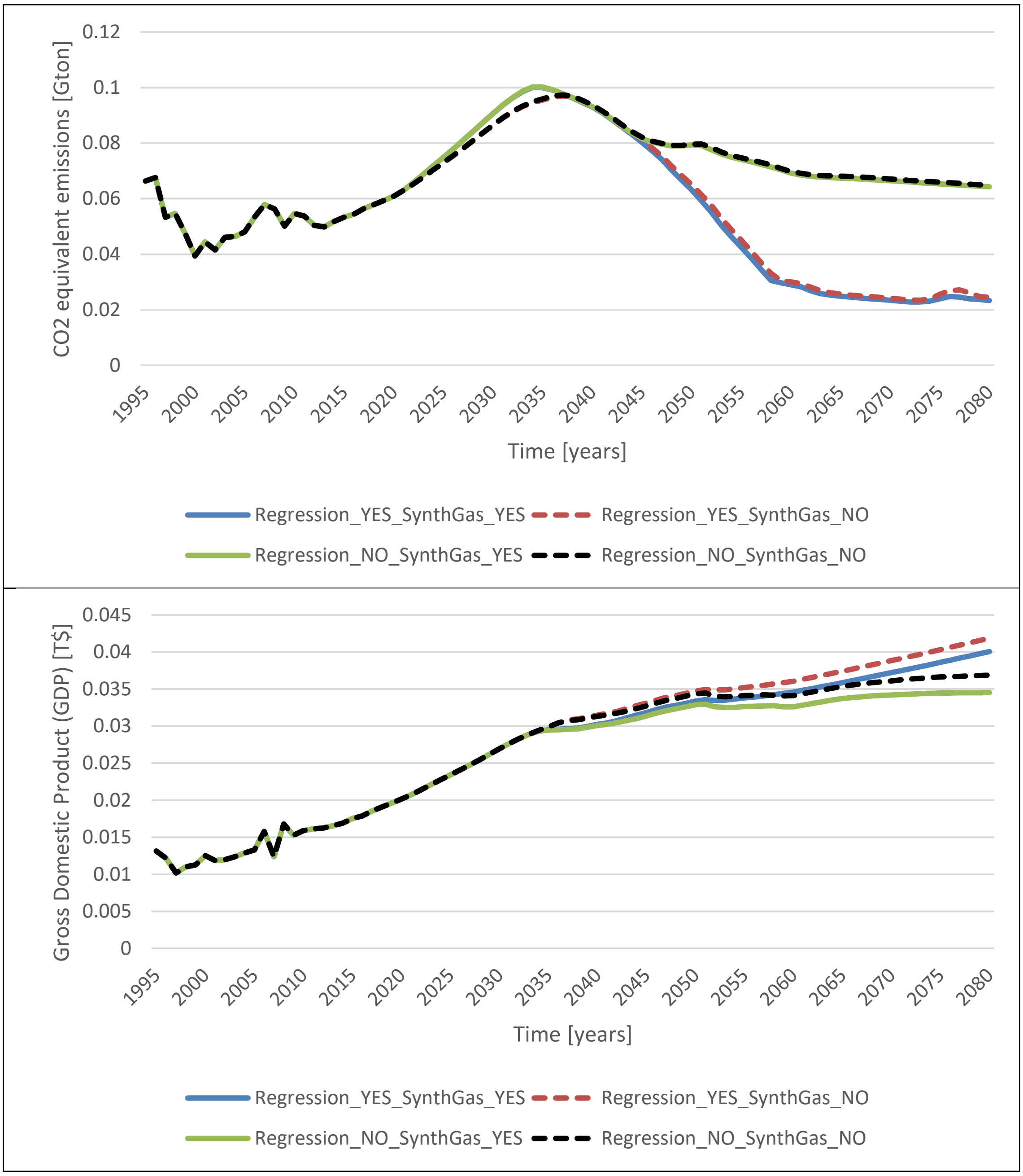



Figure 14. Results in MEDEAS-BGR for the share of renewable energy in the electricity mix (first), share of renewables in the primary energy supply (second), equivalent carbon emissions (third), and GDP (fourth). Four cases of the same scenario were represented: original method without synthetic gas (black dotted line), original method with synthetic gas (green continuous line), regression method without synthetic gas (red dotted line), and regression method with synthetic gas (blue continuous line).

Looking for information in general indicators of interest for this research, a broader scope of the new approach got value added. Regressions does not restrict as much as the old method of MEDEAS-BGR, achieving about +34% of share in the power system in 2050 (goal of 100% renewable power system was completed in 2060) and +30% of share in primary energy supply in 2050 (close to 80% at the end of simulation). As one expected, greenhouse gas emissions fall to levels below first year of simulation, 1995. Finally, behaviour of regressions showed to be also beneficial for GDP of Bulgaria because of the availability of energy to satisfy the economic demand. Among simulations, without production of synthetic gas was the best in this way. The reason was related to EROI again. A higher EROI meant less energy involved within energy system to satisfy the same demand.

## CONCLUSIONS

Along this document, we examined recent literature about traditional and novel technologies to flexibilize the power system and reach scenarios of high renewable share in different parts of the energy system, especially in the power system. It was highlighted the energy system go forward to be smart and complex in terms of sector coupling, number of different technologies involved, and uses of electricity.

This first insight provides a valuable comparison between the "old" and "new" method joining knowledge of bottom-up energy systems and top-down energy-economy-environment IAMs. A simulation of 100% renewable power system was shown based on a green growth scenario, achieving high levels of primary renewable energy in the supply side of the system. Results delivers a considerable flexibility gap between both methods, being the new one more beneficial for variable renewables in scenarios of energy transitions.

An hydrogen-based fuel was introduced in MEDEAS-BGR to realistically estimate its potential in this country, based on some industries. Results show Bulgaria has a modest potential to use hydrogen due to the weight of the related three industries in the economy. However, there was noted that this kind of fuel would reduce the EROI of the system while increase the renewable share of primary energy supply.

We would like to highlight the conclussion that reductions in LCOE of VRES technologies can provide better energy systems without the use of hydrogen as energy carrier (perhaps generation and consumption *in situ*, but not transport and distribution of the fuel).

Further research can include a higher number of industries and sectors using this fuel. Regarding with the modelling framework, it could be included as a new final energy to correctly follow the demands of this fuel and intermediate energy transformation, when data allow for get the intensity of synthetic gas by industry, as well as other possible hydrogen-based liquid fuels in transport sector.

Next steps of this approach will be conducted towards a more complete integration – soft-linking – of inputs and outputs within the modelling framework.

Finally, we would like to mention that the approach here presented is going to be implemented for 35 regions ([44], appendix A) in the WILIAM model of the Locomotion project, what opens new insights to verify the approach in different geophysical regions and fossil-fuel and renewable policy frameworks.

## REFERENCES


[1] D. Süsser, A. Ceglarz, H. Gaschnig, V. Stavrakas, A. Flamos, and J. Giannakidis, G. Lilliestam, “Model-based policymaking or policy-based modelling? How energy models and energy policy interact,” *Energy Res. Soc. Sci.*, vol. In press, no. February, p. 101984, 2021, doi: 10.1016/j.erss.2021.101984.

[2] K. Araújo, “The emerging field of energy transitions: Progress, challenges, and opportunities,” *Energy Res. Soc. Sci.*, vol. 1, pp. 112–121, 2014, doi: 10.1016/j.erss.2014.03.002.

[3] J. Markard, “The next phase of the energy transition and its implications for research and policy,” *Nat. Energy*, vol. 3, no. 8, pp. 628–633, 2018, doi: 10.1038/s41560-018-0171-7.

[4] J. Wiseman, “The great energy transition of the 21st century: The 2050 Zero-Carbon World Oration,” *Energy Res. Soc. Sci.*, vol. 35, no. October 2017, pp. 227–232, 2018, doi: 10.1016/j.erss.2017.10.011.

[5] A. Pfeifer, G. Krajačić, and D. N. Ljubas D, “Increasing the integration of solar photovoltaics in energy mix on the road to low emissions energy system – Economic and environmental implications,” *Renew. Energy*, vol. 143, no. 1310–7, 2019, doi: 10.1016/j.renene.2019.05.080.

[6] H. Lund, “Renewable heating strategies and their consequences for storage and grid infrastructures comparing a smart grid to a smart energy systems approach.,” *Energy*, vol. 151, pp. 94–102, 2018, doi: 10.1016/j.energy.2018.03.010.

[7] H. Dorotić, T. Pukšec, and N. Duić, “Analysis of displacing natural gas boiler units in district heating systems by using multi-objective optimization and different taxing approaches,” *Energy Convers. Manag.*, vol. 205, no. November 2019, 2020, doi: 10.1016/j.enconman.2019.112411.

[8] D. Meha, A. Pfeifer, N. Duić, and H. Lund, “Increasing the integration of variable renewable energy in coal-based energy system using power to heat technologies: The case of Kosovo,” *Energy*, vol. 212, 2020, doi: 10.1016/j.energy.2020.118762.

[9] F. Calise, M. D. D’Accadia, C. Barletta, V. Battaglia, A. Pfeifer, and N. Duic, “Detailed modelling of the deep decarbonisation scenarios with demand response technologies in the heating and cooling sector: A case study for Italy,” *Energies*, vol. 10, no. 10, 2017, doi: 10.3390/en10101535.

[10] V. Z. Gjorgievski, N. Markovska, A. Abazi, and N. Duić, “The potential of power-to-heat demand response to improve the flexibility of the energy system: An empirical review,” *Renew. Sustain. Energy Rev.*, vol. 138, no. September 2020, 2021, doi: 10.1016/j.rser.2020.110489.

[11] I. Ridjan, B. V. Mathiesen, D. Connolly, and N. Duić, “The feasibility of synthetic fuels in renewable energy systems,” *Energy*, vol. 57, pp. 76–84, 2013, doi: 10.1016/j.energy.2013.01.046.

[12] P. D. Lund *et al.*, “Pathway analysis of a zero-emission transition in the Nordic-Baltic region,” *Energies*, vol. 12, no. 17, pp. 1–20, 2019, doi:

10.3390/en12173337.

[13] D. Groppi, D. Astiaso Garcia, G. Lo Basso, F. Cumo, and L. De Santoli, “Analysing economic and environmental sustainability related to the use of battery and hydrogen energy storages for increasing the energy independence of small islands,” *Energy Convers. Manag.*, vol. 177, no. September, pp. 64–76, 2018, doi: 10.1016/j.enconman.2018.09.063.

[14] M. A. Ancona *et al.*, “Thermal integration of a high-temperature co-electrolyzer and experimental methanator for Power-to-Gas energy storage system,” *Energy Convers. Manag.*, vol. 186, no. March, pp. 140–155, 2019, doi: 10.1016/j.enconman.2019.02.057.

[15] Y. Van Fan, R. R. Tan, and J. J. Klemeš, “A system analysis tool for sustainable biomass utilisation considering the Emissions-Cost Nexus,” *Energy Convers. Manag.*, vol. 210, no. March, p. 112701, 2020, doi: 10.1016/j.enconman.2020.112701.

[16] A. Campione, A. Cipollina, F. Calise, A. Tamburini, M. Galluzzo, and G. Micale, “Coupling electrodialysis desalination with photovoltaic and wind energy systems for energy storage: Dynamic simulations and control strategy,” *Energy Convers. Manag.*, vol. 216, no. May, p. 112940, 2020, doi: 10.1016/j.enconman.2020.112940.

[17] V. Krakowski, E. Assoumou, V. Mazauric, and N. Maïzi, “Reprint of Feasible path toward 40–100% renewable energy shares for power supply in France by 2050: A prospective analysis,” *Appl. Energy*, vol. 184, no. 2016, pp. 1529–1550, 2016, doi: 10.1016/j.apenergy.2016.11.003.

[18] M. Ram, M. Child, A. Aghahosseini, D. Bogdanov, A. Lohrmann, and C. Breyer, “A comparative analysis of electricity generation costs from renewable, fossil fuel and nuclear sources in G20 countries for the period 2015-2030,” *J. Clean. Prod.*, vol. 199, pp. 687–704, 2018, doi: 10.1016/j.jclepro.2018.07.159.

[19] F. Feijoo, W. Silva, and T. K. Das, “A computationally efficient electricity price forecasting model for real time energy markets,” *Energy Convers. Manag.*, vol. 113, pp. 27–35, 2016, doi: 10.1016/j.enconman.2016.01.043.

[20] L. Pagnier and P. Jacquod, “How fast can one overcome the paradox of the energy transition? A physico-economic model for the European power grid,” *Energy*, vol. 157, pp. 550–560, 2018, doi: 10.1016/j.energy.2018.05.185.

[21] M. McPherson, N. Johnson, and M. Strubegger, “The role of electricity storage and hydrogen technologies in enabling global low-carbon energy transitions,” *Appl. Energy*, vol. 216, no. February, pp. 649–661, 2018, doi: 10.1016/j.apenergy.2018.02.110.

[22] M. Welsch *et al.*, “Incorporating flexibility requirements into long-term energy system models - A case study on high levels of renewable electricity penetration in Ireland,” *Appl. Energy*, vol. 135, pp. 600–615, 2014, doi: 10.1016/j.apenergy.2014.08.072.

[23] I. Batas Bjelić, N. Rajaković, G. Krajačić, and N. Duić, “Two methods for

decreasing the flexibility gap in national energy systems," *Energy*, vol. 115, pp. 1701–1709, 2016, doi: 10.1016/j.energy.2016.07.151.

[24] A. Nikolaev and P. Konidari, "Development and assessment of renewable energy policy scenarios by 2030 for Bulgaria," *Renew. Energy*, vol. 111, no. 2017, pp. 792–802, 2017, doi: 10.1016/j.renene.2017.05.007.

[25] P. Konidari and D. Mavrakis, "A multi-criteria evaluation method for climate change mitigation policy instruments," *Energy Policy*, vol. 35, no. 12, pp. 6235–6257, 2007, doi: 10.1016/j.enpol.2007.07.007.

[26] R. C. Pietzcker *et al.*, "System integration of wind and solar power in integrated assessment models: A cross-model evaluation of new approaches," *Energy Econ.*, vol. 64, no. 2017, pp. 583–599, 2017, doi: 10.1016/j.eneco.2016.11.018.

[27] L. van Beek, M. Hajer, P. Pelzer, D. van Vuuren, and C. Cassen, "Anticipating futures through models: the rise of Integrated Assessment Modelling in the climate science-policy interface since 1970," *Glob. Environ. Chang.*, vol. 65, no. May, p. 102191, 2020, doi: 10.1016/j.gloenvcha.2020.102191.

[28] A. Gambhir, I. Butnar, P. H. Li, P. Smith, and N. Strachan, "A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of BECCs," *Energies*, vol. 12, no. 9, pp. 1–21, 2019, doi: 10.3390/en12091747.

[29] F. Ueckerdt *et al.*, "Representing power sector variability and the integration of variable renewables in long-term energy-economy models using residual load duration curves," *Energy*, vol. 90, pp. 1799–1814, 2015, doi: 10.1016/j.energy.2015.07.006.

[30] H. K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke, "A review of modelling tools for energy and electricity systems with large shares of variable renewables," *Renew. Sustain. Energy Rev.*, vol. 96, no. August, pp. 440–459, 2018, doi: 10.1016/j.rser.2018.08.002.

[31] J. P. Deane, A. Chiodi, M. Gargiulo, and B. P. Ó Gallachóir, "Soft-linking of a power systems model to an energy systems model," *Energy*, vol. 42, no. 1, pp. 303–312, 2012, doi: 10.1016/j.energy.2012.03.052.

[32] F. Riva, F. D. Sanvito, F. T. Tonini, E. Colombo, and F. Colombelli, "Modelling long-term electricity load demand for rural electrification planning," in *2019 IEEE Milan PowerTech*, 2019, pp. 1–6, doi: 10.1109/PTC.2019.8810727.

[33] Ministerio para la Transición Ecológica y el Reto Demográfico, "Plan Nacional Integrado de Energía y Clima 2021-2030," *Gob. España*, p. 25, 2020.

[34] I. Capellán-Pérez, C. de Castro, and L. J. Miguel González, "Dynamic Energy Return on Energy Investment (EROI) and material requirements in scenarios of global transition to renewable energies," *Energy Strateg. Rev.*, vol. 26, no. September, p. 100399, 2019, doi: 10.1016/j.esr.2019.100399.

[35] I. Capellán-Pérez *et al.*, "MEDEAS: A new modeling framework integrating global biophysical and socioeconomic constraints," *Energy Environ. Sci.*, vol. 13, no. 3, pp. 986–1017, 2020, doi: 10.1039/c9ee02627d.

[36] H. Lund and J. Z. Thellufsen, “EnergyPLAN - Advanced Energy Systems Analysis Computer Model (Document Version 15.1),” no. September, pp. 1–189, 2020.

[37] “EU energy statistical pocketbook and country datasheets | Energy.” [Online]. Available: https://ec.europa.eu/energy/data-analysis/energy-statistical-pocketbook_en. [Accessed: 21-Apr-2021].

[38] IEA, “IEA World Energy Statistics and Balances,” IEA/OECD, Paris (France), 2019.

[39] “Query Tool.” [Online]. Available: https://www.irena.org/Statistics/Download-Data. [Accessed: 21-Apr-2021].

[40] “Global Wind Atlas.” [Online]. Available: https://globalwindatlas.info/. [Accessed: 21-Apr-2021].

[41] “Global Solar Atlas.” .

[42] “Global Solar Atlas.” [Online]. Available: https://globalsolaratlas.info/map?c=44.241264,22.906494,7. [Accessed: 21-Apr-2021].

[43] W. Shen *et al.*, “A comprehensive review of variable renewable energy levelized cost of electricity,” *Renew. Sustain. Energy Rev.*, vol. 133, no. August, p. 110301, 2020, doi: 10.1016/j.rser.2020.110301.

[44] L. Society, A. N. Enhanced, M. Tool, F. O. R. The, and T. To, “Report of the common modeling framework,” vol. 1, no. 821105, 2020.

[45] S. H. Mohr, J. Wang, G. Ellem, J. Ward, and D. Giurco, “Projection of world fossil fuels by country,” *Fuel*, vol. 141, pp. 120–135, Feb. 2015, doi: 10.1016/j.fuel.2014.10.030.

[46] I. Capellán-Pérez, M. Mediavilla, C. de Castro, Ó. Carpintero, and L. J. Miguel, “Fossil fuel depletion and socio-economic scenarios: An integrated approach,” *Energy*, vol. 77, pp. 641–666, 2014, doi: 10.1016/j.energy.2014.09.063.

## SUPPLEMENTARY MATERIAL

Table SM 1. Characteristics of the inputs for the permutation process.

| Name in the Excel | Definition | Units | Values |
|---|---|---|---|
| Wind [MW] | Capacity of wind power plants in the region. | MW | 4000, 10000,20000 |
| PV [MW] | Capacity of solar-photovoltaic power plants in the region. | MW | 4000, 10000, 20000,30000 |
| Damed hydro [MW] | Capacity of Dammed Hydro power plants. | MW | 153, 2500 |
| Ppminimum [MW] | Minimum operating capacity in Power Plants (PP1/PP2 in EnergyPLAN) | MW | 0, 1600 |
| Nuclear partload [-] | Flexibility share of nuclear power plants (totally rigid = 1) | - | 0, 1 |
| Electrification and V2G share | Electrification of transport sector as share | - | 0, 0.5, 1 |
| Jet fuel | Jet fuel consumption in transport sector | TWh/year | 0, 0.1, 0.2 |
| Bio jet fuel | Jet biofuel consumption in transport sector | TWh/year | 0, 0.1, 0.2 |
| Diesel | Diesel consumption in transport sector | TWh/year | 0, 8.9, 17.8 |
| Biodiesel | Biodiesel consumption in transport sector | TWh/year | 0, 0.5, 1 |
| Petrol | Petrol consumption in transport sector | TWh/year | 0, 2.45, 4.9 |
| Biopetrol | Biopetrol consumption in transport sector | TWh/year | 0, 0.942, 1.884 |
| Natural gas | Gris gas (natural gas) consumption in transport sector | TWh/year | 0 |
| LPG | Liquified Petrol Gas consumption in transport sector | TWh/year | 0, 2.8, 5.6 |
| Electricity dump charge | Electricity demand for electric vehicles in dump charge mode | TWh/year | 0, 1 |
| Electricity smart charge | Electricity demand for electric vehicles in smart charge mode | TWh/year | 0, 5.18, 10.36 |
| Storage | Storage in electric vehicles | GWh | 0, 56, 112 |
| Charging/discharging capacity | Capacity of electric storage in the power grid | MW | 0, 9800, 19600 |
| P2H [MW] | Power-to-heat capacity (heat pumps+electric boilers) | MW | 0, 1898.4, 3796.8 |
| P2H storage [GWh] | Storage of heat for power-to-heat facilities | GWh | 0, 56.952, 113.904 |
| Battery power capacity [MW] | Capacity of batteries in the power grid | MW | 0, 2667.1235, 5334.247 |
| Battery storage capacity [GWh] | Storage of batteries in the power grid | GWh | 0, 13.3355, 26.671 |
| PHS [MW] | Capacity of pumping mode in hydropower plants | MW | 1400, 1711.3685, 2022.737 |
| PHS [GWh] | Storage in hydropower plants to the pumping mode | GWh | 9.23, 11.283, 13.336 |
| High temperature storage [MW] | Capacity of Rockbed storage dedicated to high temperature processes | MW | 0, 833.476, 166.952 |
| High temperature storage [GWh] | Storage of Rockbed facilities dedicated to high temperature processes | GWh | 0, 13.3355, 26.671 |
| Flexible demand [-] | Percentage of electricity demand that is flexible | % | 0, 25, 50 |
| Day energy flexible [TWh] | Flexible electricity demand in the day | TWh/year | 0, 3.89, 7.78 |
| Week energy flexible [TWh] | Flexible electricity demand in the week | TWh/year | 0, 2.92, 5.83 |
| Month energy flexible [TWh] | Flexible electricity demand in the month | TWh/year | 0, 2.92, 5.83 |
| Day power flexible [MW] | Flexible capacity in the demand side of the power system in the day | MW | 0, 713.3, 1426.6 |
| Week power flexible [MW] | Flexible capacity in the demand side of the power system in the week | MW | 0, 534.98, 1069.95 |
| Month power flexible [MW] | Flexible capacity in the demand side of the power system in the month | MW | 0, 534.98, 1069.95 |
| Industry decarbonization [-] | Percentage of fossil fuels in Industry | % | 0, 50, 100 |
| Natural gas in industry [TWh] | Natural gas in Industry | TWh/year | 0, 10.88, 21.76 |
| H2 in industry [TWh] | Hydrogen in Industry | TWh/year | 0, 5.44, 10.88 |
| Electricity in industry [TWh] | Electricity in Industry | TWh/year | 0, 5.44, 10.88 |
| Synthetic gas [TWh] | Synthetic gas production | TWh/year | 0, 5, 10 |

| CONSTANT VALUES | | | |
|---|---|---|---|
| River hydro | Capacity of Run-of-River hydropower plants. | MW | 800 |
| Nuclear | Capacity of Nuclear power plants. | MW | 2000 |
| PP1 | Capacity of back-up (traditional fossil fuels) power plants. | MW | 4000 |
| CHP | Capacity of Combined Heat and Power plants. | MW | 1464 |
| District heating in gr3 | Demand of district heating in group 3. | TWh | 12.55 |
| District heating in gr2 | Demand of district heating in group 2 | TWh | 0.52 |
| Natural gas in HH | Demand of natural gas in households. | TWh | 1.89 |
| Biomass in HH | Demand of biomass in households. | TWh | 9.55 |
| Fuels in power plants and boilers | Fuel distribution. Biomass and natural gas may be replaced by synthetic gas in case hydrogen is considered a flexibility option. | - | 50:50 |

Table SM 2. First part of the Python code with which permutations are done.

```
from tkinter import filedialog
from tkinter import Tk
from tkinter import *
import openpyxl
root = Tk()
root.input = filedialog.askopenfilename(filetypes=((".xlsx", "*.xlsx"), ("All files", "*.*")),
                                        title="Open input data table")
root.withdraw()
folder_EnergyPLAN = filedialog.askdirectory(title="Open EnergyPLAN folder")

'-------------------define input file---------------------------'
path = ((root.input))
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active
m_row = sheet_obj.max_row
m_col = sheet_obj.max_column
'------------------define file storage location------------------------------'
for j in range(1, m_col+1):
    name = sheet_obj.cell(row=2200, column=j).value
    outputFile = open(r'{0}\energyPlan Data\Data\{1}.txt'.format(folder_EnergyPLAN,name), 'w')
    for i in range(1, m_row + 1):
        cell_obj = sheet_obj.cell(row=i, column=j)
        print(cell_obj.value, file=outputFile)
    outputFile.close()
```

Table SM 3. Second part of the Python code with which permutations are done.

```
'------------------import plugins----------------------'
from tkinter import filedialog
from tkinter import Tk
from tkinter import *
import os
import subprocess
import openpyxl
import pyautogui
```

```
import time
'-------------------define file locations-----------------------'
root = Tk()

folder_EnergyPLAN = filedialog.askdirectory(title = "Open EnergyPLAN folder")
folder_csv_xlsx = filedialog.askdirectory(title = "csv folder")
outputtable = filedialog.askopenfilename( filetypes = ( (".xlsx", "*.xlsx"), ("All files",
"*.*") ),title = "Open case name table" )

'-------------------run simulations in EnergyPLAN-----------------------'


path = (outputtable)
wb_obj = openpyxl.load_workbook(path)
sheet_obj = wb_obj.active
m_row = sheet_obj.max_row
m_col = sheet_obj.max_column
time.sleep(10)

for j in range (1, m_col+1):
    name = sheet_obj.cell(row=1, column=j).value
    pyautogui.click(119, 96)
    pyautogui.typewrite('{}.txt'.format(name))
    time.sleep(0.05)
    pyautogui.typewrite(['enter'])
    t1= time.time()
    pyautogui.moveTo(317, 119)
    time.sleep(0.05)
    while pyautogui.pixel(360, 119)[2] != 69:
        time.sleep(0.05)
        t2 = time.time()
        if pyautogui.pixel(360, 119)[2] == 69:
            break
        if t2-t1 > 10:
            pyautogui.typewrite(['enter'])
            pyautogui.typewrite(['enter'])
    pyautogui.click(317, 119)
    time.sleep(0.05)
    while pyautogui.pixel(360, 119)[2] != 69:
        time.sleep(0.05)
        if pyautogui.pixel(360, 119)[2] == 69:
            break
    time.sleep(0.4)
    POWERSHELL_COMMAND =
r'C:\WINDOWS\system32\WindowsPowerShell\v1.0\powershell.exe'
    subprocess.Popen([POWERSHELL_COMMAND,
                     'Get-clipboard > {0}\{1}.csv'.format(folder_csv_xlsx, name)],
                    stdout = subprocess.PIPE,
                    stderr = subprocess.PIPE)
    time.sleep(0.8)
```

```
    os.system('cmd /c "echo off | clip"')
    time.sleep(0.1)

for j in range(1, m_col + 1):
    name = sheet_obj.cell(row=1, column=j).value
    from pathlib import Path
    file = Path() / (r'{0}\{1}.csv'.format(folder_csv_xlsx, name))  # or Path('./doc.txt')
    size = file.stat().st_size
    if size < 70000:
        pyautogui.click(119, 96)
        pyautogui.typewrite('{}.txt'.format(name))
        pyautogui.typewrite(['enter'])
        pyautogui.moveTo(317, 119)
        time.sleep(0.1)
        while pyautogui.pixel(360, 119)[2] != 69:
            time.sleep(0.1)
            if pyautogui.pixel(360, 119)[2] == 69:
                break
        pyautogui.click(317, 119)
        time.sleep(0.1)
        while pyautogui.pixel(360, 119)[2] != 69:
            time.sleep(0.1)
            if pyautogui.pixel(360, 119)[2] == 69:
                break
        time.sleep(2.5)
        POWERSHELL_COMMAND =
r'C:\WINDOWS\system32\WindowsPowerShell\v1.0\powershell.exe'
        subprocess.Popen([POWERSHELL_COMMAND,
                          'Get-clipboard > {0}\{1}.csv'.format(folder_csv_xlsx, name)],
                         stdout=subprocess.PIPE,
                         stderr=subprocess.PIPE)
        time.sleep(2.5)
        os.system('cmd /c "echo off | clip"')
        time.sleep(0.5)
    else: print(name, 'is ok')
```

Table SM 4. Matlab code to generate the information of interest of  linear regression models.

```
%% Import data from spreadsheet
% Regression models for Bulgaria
%
clear, clc
%% Setup the Import Options and import the data
opts = spreadsheetImportOptions("NumVariables", 19);


% Specify sheet and range
opts.Sheet = "Regression";
opts.DataRange = "A2:S34993";
```

```
% Specify column names and types
opts.VariableNames = ["Wind", "SolarPV", "DamHydro", "Backing",
"ElectTransport", "P2H", "Storage", "FlexibleDemand", "FossilIndustry",
"SynthGas", "VarCFwind", "VarCFsolarPV", "VarCFRoR", "VarCFDamHydro",
"VarCFPP", "VarCFCHP", "VarCFNuclear", "HeatLossesStorage", "LossesV2G"];
opts.VariableTypes = ["double", "double", "double", "double", "double", "double",
"double", "double", "double", "double", "double", "double", "double", "double",
"double", "double", "double", "double", "double"];

% Import the data
currentFolder = pwd;
pathfile = strcat(currentFolder,'\Bulgaria.xlsx');
tbl = readtable(pathfile, opts, "UseExcel", false);

%% Clear temporary variables
clear opts


%% Convert to output type
Wind = tbl.Wind;
SolarPV = tbl.SolarPV;
DamHydro = tbl.DamHydro;
Backing = tbl.Backing;
ElectTransport = tbl.ElectTransport;
P2H = tbl.P2H;
Storage = tbl.Storage;
FlexibleDemand = tbl.FlexibleDemand;
FossilIndustry = tbl.FossilIndustry;
SynthGas = tbl.SynthGas;

VarCFwind = tbl.VarCFwind;
VarCFsolarPV = tbl.VarCFsolarPV;
VarCFRoR = tbl.VarCFRoR;
VarCFDamHydro = tbl.VarCFDamHydro;
VarCFPP = tbl.VarCFPP;
VarCFCHP = tbl.VarCFCHP;
VarCFNuclear = tbl.VarCFNuclear;
HeatLossesStorage = tbl.HeatLossesStorage;
LossesV2G = tbl.LossesV2G;

name_variables = tbl.Properties.VariableNames;

%% Clear temporary variables
clear opts tbl

%% Automatic Linear regression model --> VarCFwind
X_all = [Wind, SolarPV, DamHydro, Backing, ElectTransport, P2H, Storage,
FlexibleDemand, FossilIndustry, SynthGas];
criteria_value = 0.05; % To control the while loop (criteria: when all the correlation
values are below 0.05, brake the loop.
```

```
iter_y = 1; % to control the name of the sheet in the Excel file

for output = [VarCFwind, VarCFsolarPV, VarCFRoR, VarCFDamHydro, VarCFPP,
VarCFCHP, VarCFNuclear, HeatLossesStorage, LossesV2G]
   size_inputs = size(X_all);
   max_val_R = 1;
   A_corr = [Wind, SolarPV, DamHydro, Backing, ElectTransport, P2H, Storage,
FlexibleDemand, FossilIndustry, SynthGas, output];

   Intercept = 0; % Independent term of the equation in the regression model
   dependent_factors = zeros(1,size_inputs(2)); % To save the dependent terms of the
regression model, one by input took into account
   n = 1;
   names_result = {}; % To save the names of the inputs selected in correct order

   while max_val_R > criteria_value
      % Matrix of correlation (R)
      R = corrcoef(A_corr);
      max_val_R = max(abs(R(size_inputs(2)+1,1:size_inputs(2))));
      [row_max,c_max] = find(abs(R) == max_val_R);
      names_result(n,1) = name_variables(row_max(2));

      % Regression model
      mdl = fitlm(X_all(:,row_max(2)), output);
      Residuals = table2array(mdl.Residuals(:,1)); % Raw Residuals are selected
      Intercept = Intercept + mdl.Coefficients.Estimate(1);
      dependent_factors(n) = mdl.Coefficients.Estimate(2);

      % New correlation matrix and update variables
      A_corr = [Wind, SolarPV, DamHydro, Backing, ElectTransport, P2H, Storage,
FlexibleDemand, FossilIndustry, SynthGas, Residuals];
      output = Residuals;
      R = corrcoef(A_corr);
      max_val_R = max(abs(R(size_inputs(2)+1,1:size_inputs(2))));
      n = n+1;
   end

   % Write data in the Excel file
   filename = 'Bulgaria_MatlabResults_Autom.xlsx';

   Sheet_name = char(name_variables(size_inputs(2)+iter_y));
   xlswrite(filename, {'Independent factor'}, Sheet_name,'A1')
   xlswrite(filename, Intercept, Sheet_name,'B1')

   TableResults = table(names_result,
transpose(dependent_factors(1:size(names_result))));
   writetable(TableResults, filename, 'Sheet', Sheet_name, 'Range', 'A3')
   iter_y = iter_y + 1;

   % Save figure
```

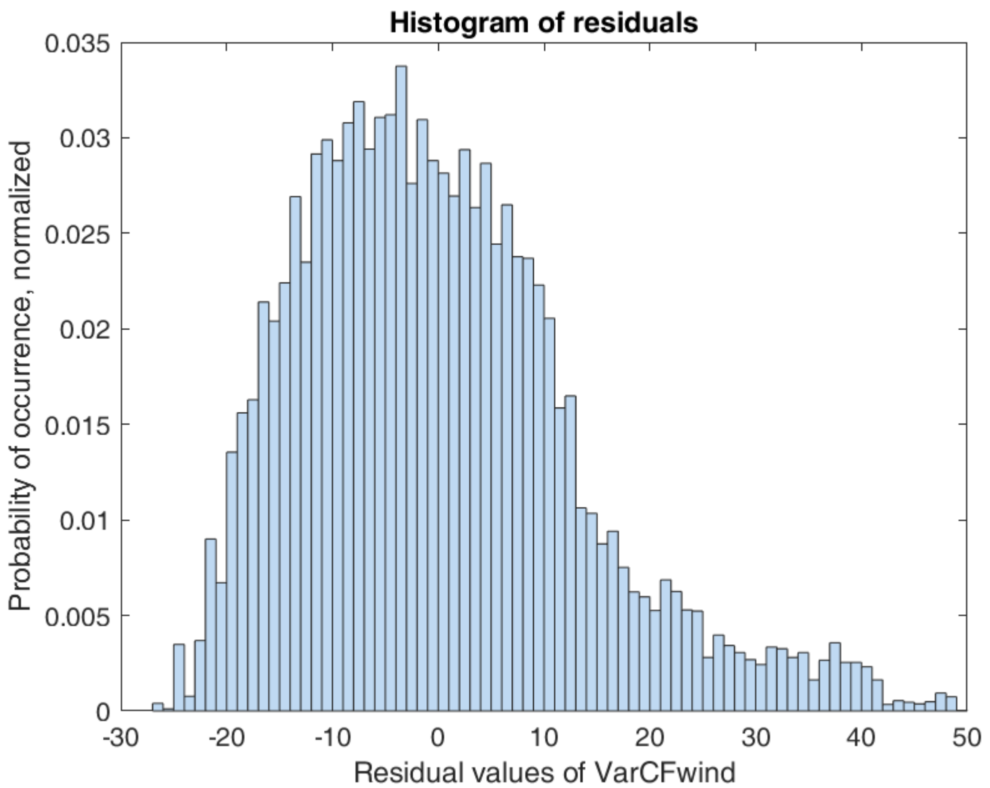
```
    fig = plotResiduals(mdl);
    xlabel(strcat('Residuals of' ,{' '}, Sheet_name));
    ylabel('Probability of occurrence, normalized')  % The area of each bar is the relative number of observations. The sum of the bar areas is equal to 1.
    saveas(fig, Sheet_name, 'png');
end
```

Table SM 5. Relevant parameters defining the green-growth scenario introduced in MEDEAS-BGR.

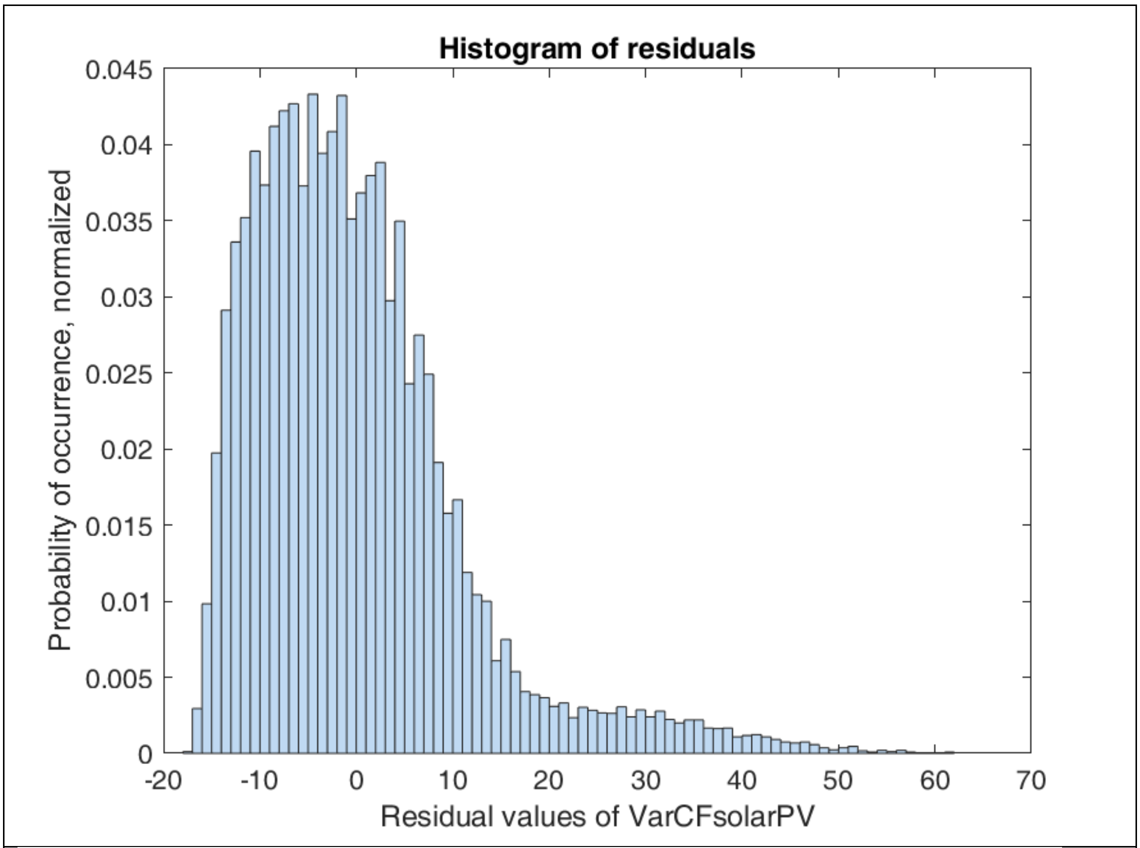
| Scenario inputs & assumptions | Value |
|---|---|
| Desired GDPpc growth (2015-2060 yearly average) | 3 %/year |
| Population growth (2015-2060 yearly average) | -0.59 %/year |
| Target labour share 2050 | 50 % |
| A matrix | Constant values of 2009 IOT. |
| Phase-out oil for electricity and heat? | No, constant current share. |
| Efficiency improvements (Final energy intensity) | 2x times increase historical efficiency improvement trends by sector/households and fuel. |
| **Inland and households transport. Electric vehicles&hybrid shares target per category in 2050.** | |
| 4-wheel vehicles (including light cargo) | 60 % |
| 2-wheel vehicles | 92.54 % |
| Heavy vehicles | 20 % |
| Bus | 90 % |
| Train | 95 % |
| Nuclear installed capacity | No more nuclear capacity is installed. Current capacity is depreciated over the simulation. |
| Recycling rates of minerals (19 minerals) | 5 %/year of improvement in the rate up to maximum. |
| **Renewables** | |
| **Potential of installed capacity of power plants** | |
| Hydropower | 17.52 TWh |
| Geothermal | 4.38 TWh |
| Pumped Hydropower Storage | 4.38 TWh |
| Wind onshore | 21.9 TWh |
| Wind offshore | 21.9 TWh |
| Oceanic | 0.438 TWh |
| Solar-CSP | 0 TWh |
| **Annual capacity growth of RES for electricity** | |
| Hydropower | 1.4 %/year |
| Geothermal | 6.8 %/year |
| Solar-PV | 19 %/year |
| Solar-Thermal (for heating) | 14 %/year |
| Solar-CSP | 7.2 %/year |
| Average solar irradiation | 1300 $kW/m^2$ |
| Wind onshore | 17.4 %/year |
| Wind offshore | 25.4 %/year |
| Oceanic | 0.8 %/year |
| Solid bioenergy | 7 %/year |
| $2^{nd}$ Gen cropland | 8 %/year |
| $3^{rd}$ Gen cropland (starting 2025) | 8 %/year |
| Residues for non-biofuels (starting 2020) | 20 %/year |
| Biogas | 30 %/year |

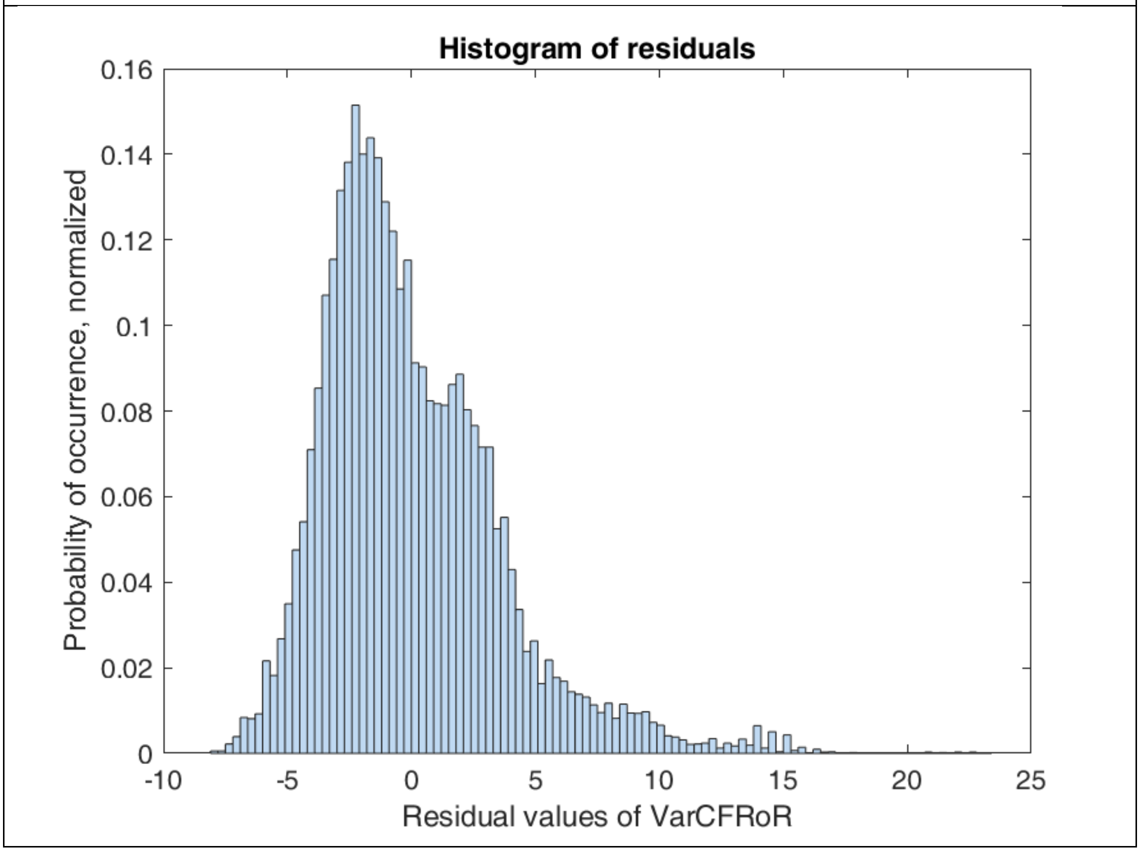
| Non-renewable energies depletion curves** | |
|---|---|
| Conventional oil | Mohr15 high-EU [45] |
| Unconventional oil | Mohr15 Low-EU [45] |
| Conventional gas | Mohr15 BG-EU [45] |
| Unconventional gas | Mohr15 Low-EU [45] |
| Coal | Mohr15 BG-EU [45] |
| Uranium | EU domestic uranium extraction 2015 |
| Climate Change impacts | No activated |
| **Power-to-Gas** | |
| Synthetic gas demand | Ramp from 0% to 100% of gases demand of three industries, from 2021 to 2041. |
| Capacity of Power-to-Gas facilities | Ramp of slope 0.2 MW/year from 2021 to 2041 |
| Capacity factor of Power-to-Gas facilities | 0.95 |
| **Thermal power plants: time of planning [years] / time of construction [years] / lifetime [years] / heat-power ratio** | |
| Steam power turbines | 1 / 2 / 40 / 4.15 |
| Combined-cycle gas turbines | 1 / 2 / 40 / 4.15 |
| Diesel internal engines | 1 / 2 / 40 / 4.15 |
| Required capacity of thermal power plants | 50 MW/year for the three categories, from 2018 to 2050. |

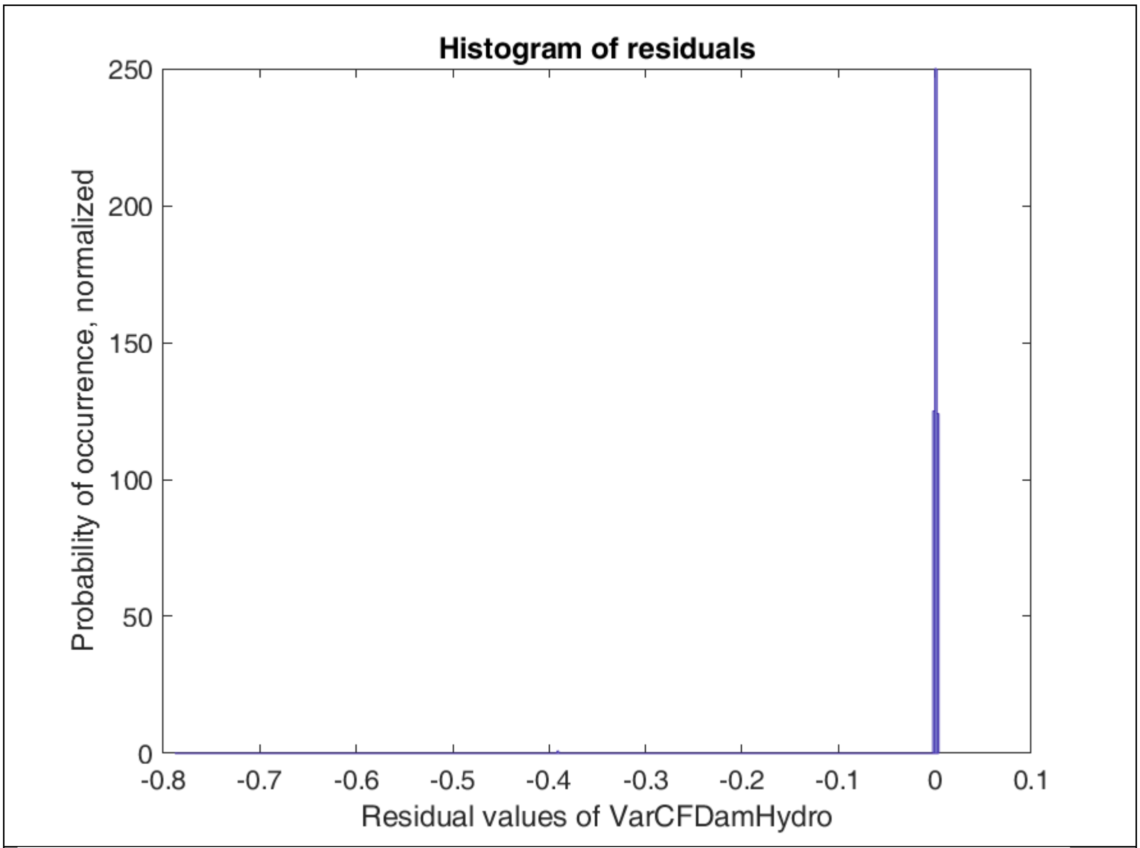
Table SM 6.probability distributions of the raw residuals by linear regression model used in the article can be shown below.

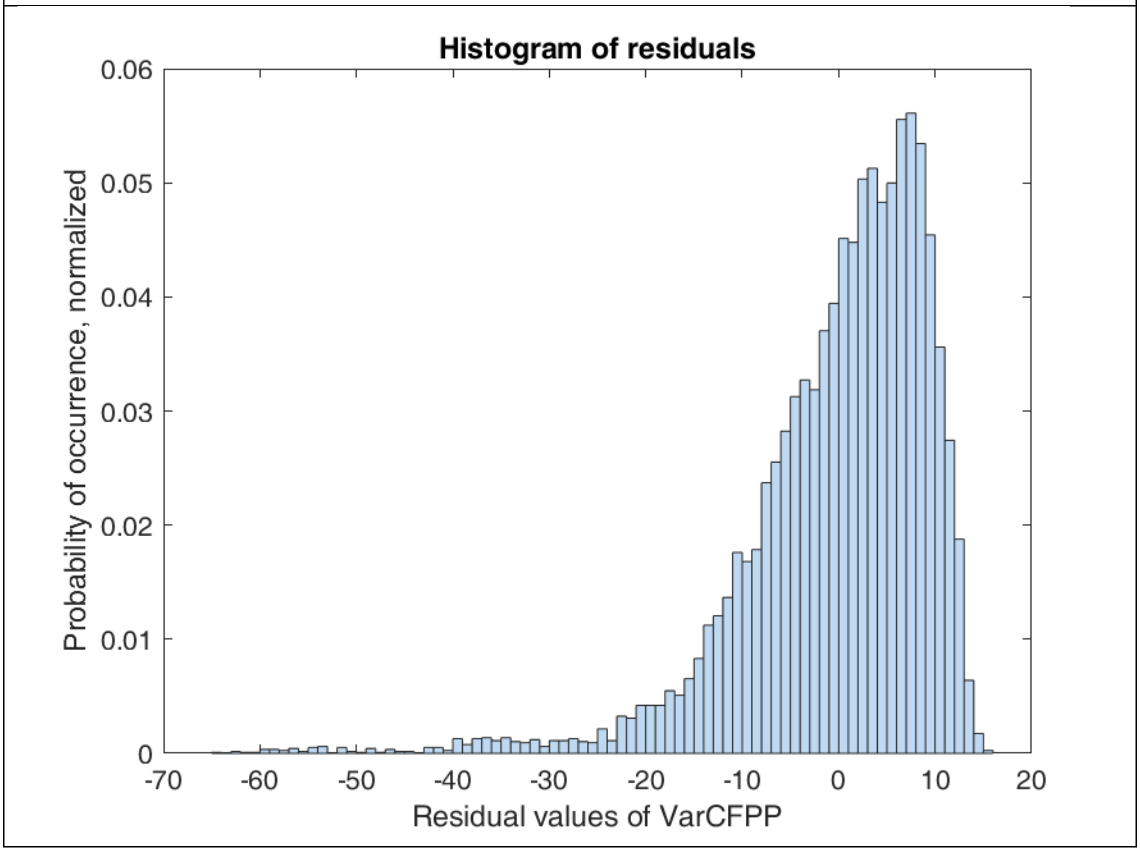


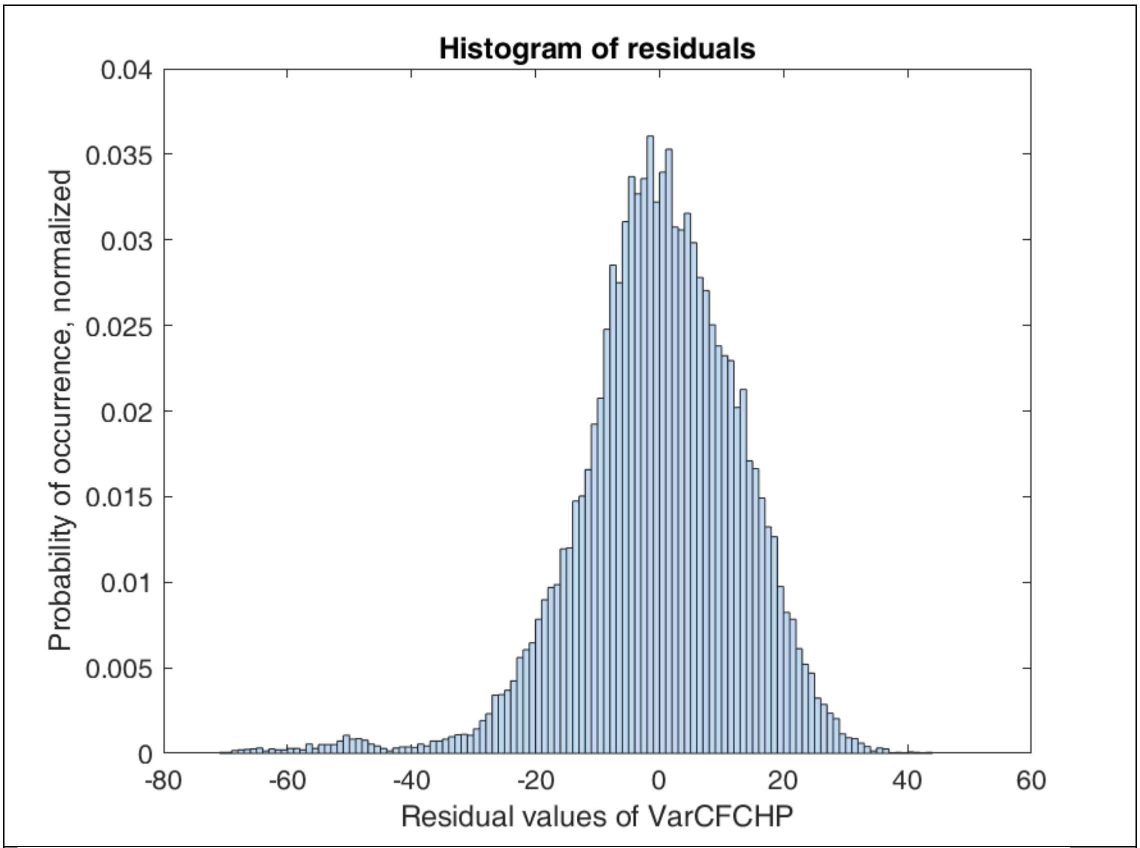

** The methodology to build the depletion curves from time series data from Mohr et al is documented in [46].

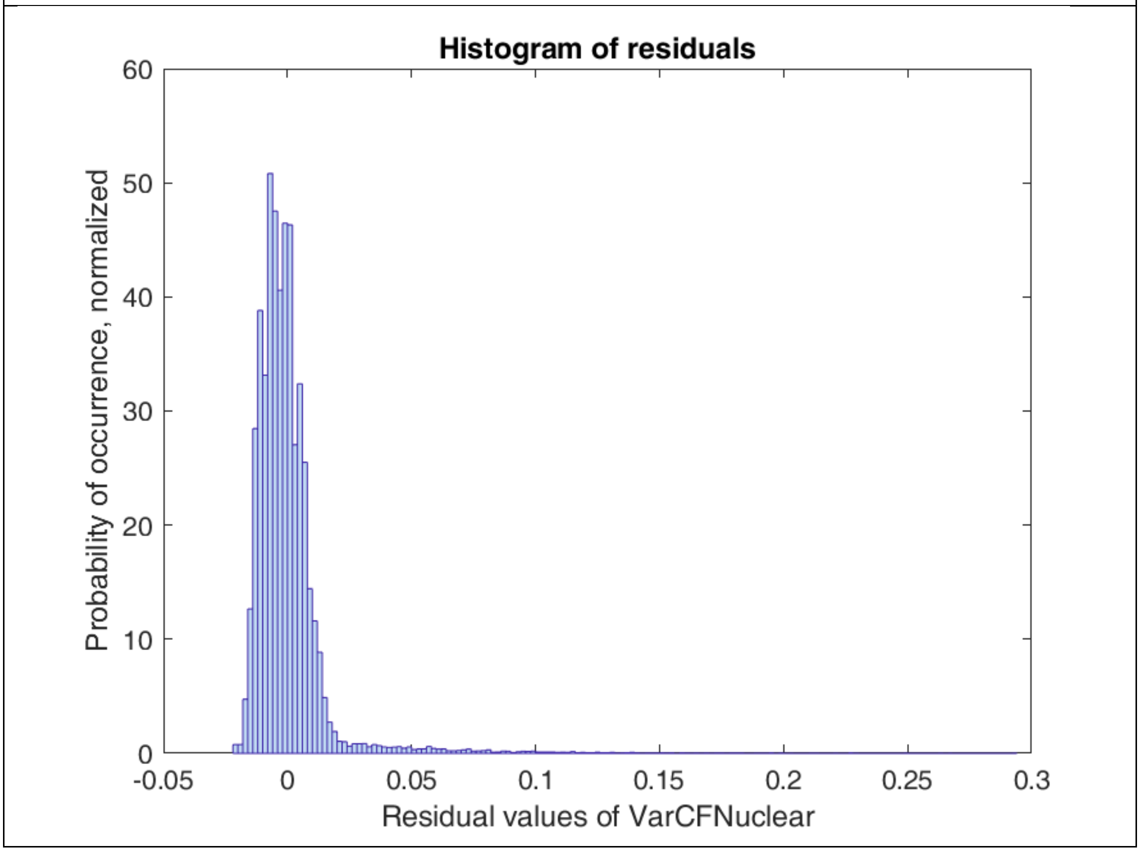

Histogram of residuals
Probability of occurrence, normalized
Residual values of VarCFsolarPV
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-20
-10
0
10
20
30
40
50
60
70


Histogram of residuals
Probability of occurrence, normalized
Residual values of VarCFRoR
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-10
-5
0
5
10
15
20
25


Histogram of residuals
Probability of occurrence, normalized
Residual values of VarCFDamHydro
250
200
150
100
50
0
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1


Histogram of residuals
Probability of occurrence, normalized
Residual values of VarCFPP
0.06
0.05
0.04
0.03
0.02
0.01
0
-70
-60
-50
-40
-30
-20
-10
0
10
20


Histogram of residuals
Probability of occurrence, normalized
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-80
-60
-40
-20
0
20
40
60
Residual values of VarCFCHP


Histogram of residuals
Probability of occurrence, normalized
60
50
40
30
20
10
0
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
Residual values of VarCFNuclear