MARIA GREITZER

ENERGY FORECASTING. FOCUS: NATURAL GAS

DISSERTATION HAFENCITY UNIVERSITY HAMBURG

A dissertation submitted to HafenCity University Hamburg in fulfilment of the requirements of the "Promotionsordnung der HafenCity Universität Hamburg" for the degree of Doktor-Ingenieur (Dr.-Ing.)

Maria Greitzer

Supervisor Prof. Dr.-Ing. Ingo Weidlich, HafenCity University

Second Supervisor Prof. Michael Bradshaw, The University of Warwick

TUFTE-LATEX.GITHUB.IO/TUFTE-LATEX/

First printing, April 2021

Contents

List of Figures

List of Tables

Acknowledgements

I OWE MY DEEPEST GRATITUDE to Prof. Dr.-Ing. Ingo Weidlich of HafenCity University. He was always willing to let find my own path. I have learned a tremendous amount from him.

I would also like to thank Prof. Michael Bradshaw of the University of Warwick for being my supervisor. We share a professional interest in the natural gas sector and geopolitics.

FOR THE MODELLING SECTION, special thanks go to Suli Li (Cornell Tech University) and Lindsey Nield (Colorado School of Mines), who taught me a lot about the world of statistics and data science. Thank you to the RISE Germany program for enabling this cooperation.

FOR THEORETICAL SECTION 3.1, I would like to thank Comfort Mosha for an unforgettable Friday afternoon (the 22nd February 2019) spent discussing the visualization of our ideas and Xhelona Haveriku for exploring analogies between music and the natural gas sector. Special thanks go to Jenna Arnold for ensuring the section's language accuracy.

I WOULD LIKE TO THANK HAFENCITY UNIVERSITY for enabling me to teach my courses "Energy in the International Context", "Physical Principles through stories" as well as the REAP course "Urban Energy Flows". Time spent with the students attending these lectures has been the best time of my life here in Hamburg.

THANKS TO (IN ALPHABETICAL ORDER):

- Nicolas Brodu (tenured member of Geostat, Geometry & Statistics in acquisition data) for his advice regarding complexity,
- Ivan Dochev of HafenCity University for our talks on the modelling component,
- Alice Brüssel-Kurbanov of HafenCity University for the management of procedural issues regarding the doctoral study, and for all the supporting words during research,
- Pakdad Langroudi of HafenCity University for helping me use LaTeX,
- Rajeev Rajaram (Department of Mathematical Sciences, Kent State University, Kent, OH, US) for his work on measuring diversity of variables in a data set,

- David Szeplaky, Slovak Gas Industry (SPP), for his expert advice on technical aspects of a gas pipeline (Ukraine-Slovakia-Austria).

  I would like to thank my brother Michal for helping me with the visual side of this work, my mum, Maria and my dad, Michal for listening and their optimism.

Thank you.

Maria Greitzer

Hamburg, Germany

January 10th, 2021

Summary

THIS WORK EXPLORES the possibilities of forecasting German gas imports and German gas consumption in detail by:

- Identifying proper terminology, with initial thoughts on the status quo of forecasting, using analogies with music (section 3.1),
- Understanding forecasting models based on statistics and machine learning (section 3.2),
- Applying domain knowledge while constructing data sets (data acquisition)[1] - various models have been tested on low resolution data sets (monthly and yearly data). In classical forecasting, forecasts are conducted at company level (e.g. electricity load forecasting) for a short-term period fulfilling the requirement of profit maximization. In contrast, this work's forecasts spatially cover 1) all import nodes into Germany, 2) gas consumption in Germany with the focus on *energy security*.
- Examining various notions of complexity in relation to forecasting

[1] In this research setup, forecasting with real-world data sets serves as a tool for technical infrastructure management, similarly to business forecasting. Pursuing forecasting methods forward would require standard data sets known from forecasting competitions (e.g. Kaggle, M4).

Novelty in methodology

- Considering infrastructure as the upper limit for the forecasting model (section 4.1)
- Testing models for ex post forecasting of gas imports into Germany (section 4.2)
- Testing models for *ex post* forecasting of gas consumption in Germany (section 4.3). Producing an ex ante forecast by answering *"what will be the future gas consumption in Germany in next ten years"* would cause the forecasting error to increase dramatically due to the uncertainty of future input values, such as population (lower risk) or, weather (higher uncertainty), which would lead to prediction intervals being too wide to make any statement about the future imports or consumption.
- Applying *complexity* measures such as approximate entropy ApEn and sample entropy to two self-constructed real-world data sets on gas imports to Germany and gas consumption in Germany (section 3.3)
- Section 3.1 - (a previously published conference paper) represents the creative part of this work carried out at the beginning of the

research in 2018-2019. This component suggests analogies between forecasting and music.

Zusammenfassung

DIESE ARBEIT UNTERSUCHT die Möglichkeiten zur Prognose von Deutschen Gasimporten sowie des Gasverbrauchs in Deutschland im Detail, wie folgt:

- Identifikation klarer Terminologie, mit ersten Gedanken zum Status Quo der Prognosetechnik (Abschnitt 3.1)
- Erweiterung des Verständnisses von Prognosemodellen, die aus den Themenbereichen der Statistik und des maschinellen Lernens stammen (Abschnitt 3.2)
- Anwendung von Fachwissen bei der Konstruktion von Datensätzen (Data acquisition)[2] - verschiedene Modelle wurden auf niedrig aufgelösten Daten getestet. Bei der klassischen Vorgehensweise werden Prognosen auf Unternehmensebene (z.B. Stromlastprognose) für einen kurzfristigen Zeitraum durchgeführt, der die Anforderung der Gewinnmaximierung erfüllt. Im Gegensatz dazu umfassen in diesem Werk präsentierte Prognosen räumlich 1) alle Importknotenpunkte nach Deutschland und 2) den Gasverbrauch in Deutschland mit dem Fokus auf die Energiesicherheit.
- Untersuchung verschiedener Begriffe von Komplexität in Bezug auf Prognosen

[2] In diesem Setup wurde die Vorhersage mit Realwelt-Datensätzen (Monats- und Jahresdaten) vorgenommen. Realwelt-Datensätze dienen als Werkzeug für die Prognoseerstellung im Bereich Technisches Infrastrukturmanagement, ähnlich wie bei Geschäftsprognosen. Um die Prognosemethoden weiterzuentwickeln, werden Standarddatensätze benötigt, die aus Prognosewettbewerben bekannt sind z.B. Kaggle, M4.

Neuerungen in der Methodik

- Betrachtung der Infrastrukturkapazität als Obergrenze für das Prognosemodell (Abschnitt 4.1)
- Testen von Modellen für die Ex-post-Prognose von Gasimporten nach Deutschland (Abschnitt 4.2)
- Testen von Modellen für die Ex-post-Prognose des Gasverbrauchs in Deutschland (Abschnitt 4.3). Erstellung von Ex-Ante-Prognosen durch Beantwortung der Frage: "Wie hoch wird der zukünftige Gasverbrauch in Deutschland in den nächsten zehn Jahren sein?" würde dazu führen, dass der Prognosefehler aufgrund der Unsicherheit der zukünftigen Werte von Inputs wie Bevölkerung (geringeres Risiko) oder Wetter (höhere Unsicherheit) dramatisch ansteigt. Dadurch würden die Vorhersageintervalle zu groß, um eine zutreffende Aussage über die zukünftigen Importe oder den Verbrauch zu treffen.

- Die Anwendung von Komplexitätsmaßen wie der ungefähren Entropie (ApEn) und der Probenentropie auf zwei selbst konstruierte Realwelt-Datensätze zu Gasimporten nach Deutschland und Gasverbrauch in Deutschland (Abschnitt 3.3)

Abschnitt 3.1 steht für den kreativen Teil der Arbeit zu Beginn der Forschung (2018-2019) und erstellt Analogien zwischen Prognosen und Musik.

1 Preface

While writing, statements and expressions with lower information value were left out to keep this thesis compact with respect to the reader's time.

My personal aim was to continue with this research topic only if it still interested me. To my surprise, I enjoyed the process until the last typed sentence. This one.

2 Introduction

GAS FORECASTING[1] covers the forecasting of gas demand of a city or a country, gas prices, and gas imports. The history of forecasting methods starts with classical mathematical relationship models based on statistics (e.g.linear regression, autoregressive integrated moving average) and ranges to artificial intelligence (e.g. artificial neural network (ANN), deep neural networks). The overview of the latter methods is provided by Merkel et al. (2018),[2] whereas Tamba et al. (2018)[3] conducted a literature survey on gas forecasting. Literature on forecasting methods is immense. For grey system theory in China alone, more than 50,000 papers were retrieved from academic periodical databases covering the period from 1982 (the date of the first publication on this topic) to 2010 (Liu (2011)).[4] Therefore, instead of conducting another literature survey, this work captures the essential idea of the most common gas forecasting methods along with examples from literature.

[1] In further statements, we refer to gas only, omitting an adjective *natural*.

[2] Merkel, G., Povinelli, R., and Brown, R. (2018). Short term load forecasting of natural gas with deep neural network regression. *Energies*, 11(8)

[3] Tamba, J. G., Essiane, S. N., Sapnken, E. F., Koffi, F. D., Nsouandélé, J. L., Soldo, B., and Njomo Donatien (2018). Forecasting natural gas: A literature survey. *International Journal of Energy Economics and Policy*, (8(3)):216–249

[4] Liu, S., editor (2011). *Proceedings of 2011 IEEE International Conference on Grey Systems and Intelligent Services (GSIS) with the 15th WOSC International Congress on Cybernetics and Systems, Nanjing, China, 15 - 18 September 2011*, Piscataway, NJ. IEEE

THE THEORY CHAPTER 1) examines energy forecasting and energy modelling, 2) provides theoretical background to understand the methods used in the second section, and 3) discusses various notions of complexity and their eventual use for forecasting. Alternative approaches to the topic would cover: 1) earth science and perspectives on climate change and energy scarcity; 2) a purely economic perspective on imports and exports of energy resources, as well as incentives and sanctions of governments; and 3) energy security perspective from an International Relations (IR) point of view.

THE EXPERIMENTAL CHAPTER applies models to self-constructed data sets, and forecasts gas imports into Germany and gas consumption at the country level. Models are understood as tools used in energy engineering and policy making but they have not become the object of the research per se. If this work intended to introduce new methods for energy forecasting, this work would use standard data sets which would enable the comparison of the models' performance.

2.1 Research questions

- RQ1. What are current approaches to predict gas imports and gas consumption at the country level? What is the impact of the domain knowledge of the gas sector on forecasting?

  GAS CONSUMPTION AT COUNTRY LEVEL depends on the amount of gas used to cover residential heat demand as well as on the energy intensiveness of industry. Models may include this type of domain knowledge or work with time series methods. As residential heat demand is correlated with weather, the accuracy of weather forecasts influences the accuracy of forecasting gas load (Franco and Fantozzi (2015)).[5]

  GAS IMPORTS, on the other hand, are more related to contracts among exporting and importing countries and, in the long run, on the environmental policies of gas use in future decades.

  Regarding current approaches, the forecasting community compares machine learning methods[6] with statistical methods[7] in terms of accuracy criteria. This work does not contribute to this debate as both terms "statistical" and "machine learning" are ill-defined. Here, *structured* and *unstructured* models will be discussed.

  TO UNDERSTAND HOW FORECASTING METHODS WORK, a few models will be tested on two self-constructed data sets for German gas imports and German gas consumption. The focus lies in constructing data sets and understanding current models instead of developing their more complex versions.

- RQ2. Is there any relation between the complexity of the data set, forecasting models and the process modelled (gas flows) to the accuracy of forecasts?

  AS FOR THE COMPLEXITY OF A DATA SET, a few measures such as *the Sample Entropy* will be tested on the above-mentioned data sets. All data sets represent just a sample providing incomplete information on the process forecasted; thus there is the inherent uncertainty about including the most representative variables that influence the output.

[5] Franco, A. and Fantozzi, F. (2015). Analysis and clustering of natural gas consumption data for thermal energy use forecasting. *Journal of Physics: Conference Series*, 655:012020

[6] Examples for ML models: support vector machine models, decision trees, artificial neural networks, nonlinear programming.

[7] Examples for statistical models: autoregressive integrated moving average, linear regression, logistic regression.

There is no clear distinction between simple and complex models. The organizers of the *M*3 forecasting competition listed, among others, naïve models and exponential smoothing models as examples of simple models (Green and Armstrong (2015)). Complexity of the models will be expressed in various ways; for example in the computational time required for the model to be run, in the number of parameters[8] and degrees of freedom.

[8] Over-parametrized models will capture not only essential information from the data set but also noise (Pelikán (2014)).

Complexity of the process being modelled increases with time, therefore there are limited possibilities to forecast for a long-term horizon. The essence of the process is also crucial; forecasting the behaviour of economic systems cannot outperform the performance of forecasts for other system with chaotic properties, such as meteorological ones (Makridakis (1995)).[9]

[9] Makridakis, S. (1995). Forecasting accuracy and system complexity. *RAIRO - Operations Research - Recherche Opérationnelle*, 29(3):259–283. http://www.numdam.org/item/RO_1995__29_3_259_0/

2.2 *What is forecasting?*

Forecasting covers methods for producing forecasts by using historical data to make predictions and to determine trends. This relatively new discipline seeks to gain independence from the fields of statistics and machine learning, although the first generation of forecasters arose from statisticians and econometricians.

"A forecasting model (predictive model, autoregressive model) is a software-implemented model of a system, process, or phenomenon, usable to forecast a value, output, or outcome expected from the system, process, or phenomenon." (Baughman et al. (1988)).

In renewable energy forecasting, forecasts of solar radiation and wind speeds are used to optimize forecasts for electricity production from renewable energy sources. These natural phenomena affect output to a high degree and are the main source of uncertainty (when it comes to renewable energy production).

Predictions, meaning the outcome of the forecasting, support company decisions on daily operations[10] (e.g. fuel storage, wood pellet bunkers), strategic decisions on investments and more. Although energy forecasting is regarded as an academic sub-discipline, its potential to impact decisions in the energy sector is expected to grow.

[10] Intra-day and day-ahead forecasts assist in scheduling partially flexible conventional generation (i.e. coal power plants, gas power plants).

2.3 *Scenario modelling*

Scenario modelling is a tool for governments' decision-makers to introduce policies regarding climate change, for example. Often, a goal (e.g. the share of renewable energy in a system) is already defined and research questions relate to options for achieving the goal; e.g. how many megawatts of capacity and which type of installed plant is needed to produce *almost* all electricity from renewable energy sources in 2050? To answer such a question, techno-economic optimization models are used, for example the REMIX-Europe Model of German Aerospace Center.[11] This model optimizes capacity, applying a minimum cost function for electricity production under the condition of 100% renewable electricity.

[11] In German, Deutsches Zentrum für Luft und Raumfahrt, DLR.

UNTIL NOW, well-known modelling scenarios have underestimated the growth of renewable energy in developed countries and overestimated gas consumption in developing countries. Thus, energy policies have been based on assumptions that have not come true. Installed capacity (and electricity production) for wind and solar had to be adjusted upwards in retrospect for almost every forecasting report. This fact hardly surprises anyone, unless the boom of forecasting and modelling methods is taken into consideration. Friedrichs (2013)[12] offers a few reasons for such a development within bodies serving as authorities for estimating energy supply: the International Energy Agency (IEA), the US-Energy Information Administration (EIA), and private entities such as BP and Shell. As for the IEA, the majority of staff have a background in economics and a strong belief in the capability of market forces to reach the optimum balance of supply and demand. Hence, "*until* 2008 *the standard practice of the IEA has been to extrapolate trends in energy demand, and simply to assume that future demand will be met via the market mechanism*" (Friedrichs (2013)).

[12] Friedrichs, J. (2013). *The future is not what it used to be: Climate Change and Energy Scarcity*. MIT Press, Cambridge, Mass. ISBN 9780262019248

Scenario modelling is compared to forecasting in the theoretical section, but not further explored.

2.4 *Why complexity?*

COMPLEXITY AS A TERM stands for distinct concepts in forecasting, mathematics and information theories; the same statement can also be applied to entropy. Someone considering a career path in science spends years studying one singular concept of complexity or entropy; yet there is no reason that would force them to crosscheck their research object with other disciplines. As a result, the same concept has a different name across various disciplines, or the concepts of the same name mean different things, as will be shown for complexity and entropy. A holistic explanation of these research phenomena is yet to come, as was the case with physics around the end of the 19^{th} century.

FORECASTING MODELS HAVE BECOME INCREASINGLY COMPLEX, although accuracy has hardly increased. Still, simplicity repels and complexity lures because *"researchers are aware that they can advance their careers by writing in a complex way"* (Green and Armstrong (2015)).[13] This effort is rewarded by highly-ranked journals that favour complexity over interpretability. However, besides efforts in academia, the primary goal of forecasting is to produce forecasts that support a decision-making process.

[13] Green, K. C. and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8):1678–1685

In the modelling section, models labelled as *simple*, such as a regression, have been included to check whether the argument by Green and Armstrong (2015) that "*most simple methods are more accurate than complex methods*" remains valid in this research.

2.5 Scope

THIS WORK EXPLORES capabilities of statistical and machine learning forecasting for the energy sector and their relation to complexity. This work also aims to provide methodological and linguistic clarity on forecasting and scenario modelling, especially for the energy sector.

THE CHALLENGE OF THIS THESIS was to investigate three unrelated fields of science; understanding the research jargon of institutions such as the School of Management, or the Department of Econometrics and Business Statistics, and combining the acquired knowledge with energy engineering. For this reason, definitions have been included as margin notes.

3 Theoretical Background

3.1 Energy forecasting vs energy modelling

Introduction

Dr. Tao Hong pointed out four main issues in energy forecasting in his speech on the Global Forecasting Competition in 2014 (Hong (2014))[1]: impractical research, lack of benchmarking data, a hard-to-reproduce process, and limited educational programmes and courses. Here,[2] this research addresses the first issue - impractical research - since energy forecasters devote their efforts to describing a methodology setup, leaving the interpretation of results aside. Impractical research relates to the broader problem described back in 2000 by David J. Hand in his comment to Lindley (2001)[3]:

"The focus (in statistical journals) seems to be increasingly on narrow technical advance into increasingly specialized areas, with greater merit being awarded to work which is more abstract and more divorced from the realities of data."

First, the relationship between energy forecasting and energy modelling is discussed in theory by introducing an analogy with music and by exploring energy models applied to gas forecasting. Most energy models are rooted in econometric theory and statistics, therefore a missing piece - the social sciences - is discussed with regard to the energy model results.

Although far from perfect, forecasting models find regular use in policy making and in amending strategies to secure the energy supply of a country or supra-region, as recognised by Zhang and Yang (2015).[4] *Energy* forecasting encompasses making predictions for gas and electric load, prices, electricity generation from weather-dependent sources, etc. The term *energy* determines the use of forecasting

[1] Hong, T. (06/30/2014). Global energy forecasting competition. past, present and future. https://forecasters.org/wp-content/uploads/gravity_forms/7-2a51b93047891f1ec3608bdbd77ca58d/2014/07/HONG_TAO_ISF2014.pdf Last accessed 2020-12-20

[2] The first version of this chapter has been published in Grajcar (2019).

Grajcar, M. (2019). Energy Forecasting vs Energy Modelling. jazz improvisation vs Symphony. http://cyseni.com/archives/proceedings/Proceedings_of_CYSENI_2019.pdf Last accessed 2020-12-20

[3] Lindley, D. V. (2001). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337. https://doi.org/10.1111/1467-9884.00238

[4] Zhang, W. and Yang, J. (2015). Forecasting natural gas consumption in China by Bayesian Model Averaging. *Energy Reports*, 1:216–220. https://doi.org/10.1016/j.egyr.2015.11.001

methods in the respective field of interest only; the same method may be used for various study objects: output growth, exchange rate, or inflation, to mention a few.

To ANALYSE CURRENT TRENDS in energy forecasting and modelling in relation to the energy sector, examples were chosen according to these criteria:

- The subjects of energy forecasting and modelling are relevant for the future security of supply of a respective country or region. From this perspective, long-term forecasting[5] is of higher importance than short-term or medium-term forecasting.
- A method used represents a current trend in energy forecasting.

[5] Makridakis et al. (1998), one of the world's leading experts on forecasting, pleads for a special attitude for long-term forecasts; "*When forecasting the long-term, a less formal approach is often better. This can involve identifying and extrapolating mega trends going back in time, using analogies, and constructing scenarios to consider future possibilities.*"

DROPPING THE NECESSITY of a deep understanding of energy processes has been regarded as a great advantage of new models such as neural networks and fuzzy logic,[6] and has led to a boom in published research articles. However, most analysts view oil/gas prices or demand as just one out of many research objects to validate their forecasting methodologies – without being genuinely interested in the implications of their research for the energy sector. In line with the nature of forecasting, analysts typically spend 95% of their articles on methodology setup.

[6] Rules of fuzzy logic depend on maxima and minima, in contrast to sums and products of the rules of the probability calculus (Lindley (2001)).

ACCURACY DOES NOT HAVE TO BE THE HIGHEST GOAL when producing forecasts (Ascher (1978)[7], Overholt (2000)).[8] Ordered by public institutions, forecasts set the mood for a decade by predicting the share of *renewable energy* in the energy mix, making distinctions between *smart grids* and traditional grids and promoting e.g. *Sektorkopplung* in Germany. Expressions such as these *in italics* represent, in chronological order, trendy words that have entered the virtual space of public debates on energy. One plays around with phrases such as *econometric methods, neural networks* and *fuzzy logic* in a similar fashion in energy forecasting. The explosion of published articles has the unspoken intention of enhancing the prestige of forecasters and capturing the attention of the public or academia.

[7] Ascher, W. (1978). *Forecasting: An appraisal for policy makers and planners*. Johns Hopkins University Pr, Baltimore. ISBN 0801820359

[8] Overholt, W. H. (2000). Forecasting: An Appraisal for Policymakers and Planners. *Policy Sciences*, 33(1):101–106

3.1.1 Energy forecasting - examples

TRADITIONALLY, energy forecasting has been based on two approaches: on the extrapolation of historical data to predict the future, and on modelling (simulation). The first group of methods deals with the concept of curve fitting, e.g. using multiple regression to fit a polynomial function to a series of data points. The more polynomials that are chosen, the smaller the fitting error. However, the fitting (training) error is neither the main criterion of forecast quality, nor a good estimate of the test error.

FROM A FEW DOZEN MODELS, this work presents the easiest explanation of the logic behind the ARIMA model. Hong (2013)[9] explained the autoregressive process applied to forecasting the electric load as: *"the current value of electric load is often expressed as linear combination of previous actual electric load values and with a random noise."* In the model-fitting process, analysts try to fit their model to the data until residuals become white noise and forecasts look reasonable. Similarly, operational staff at energy production sites adjust their predictions (i.e. regression model outcomes) based on their knowledge of the plant. On a countrywide' scale this would be impossible to do.

ARIMA - Auto-Regressive Integrated Moving Average Model, presented by George Box and Gwilym Jenkins in the 1970*s*.

[9] Hong, W. S. (2013). *Intelligent energy demand forecasting*, volume 10 of *Lecture notes in energy*. Springer, London and Heidelberg and New York and Dordrecht. ISBN 9781447149682

White noise is (statistically) time reversible (i.e. invariant under time reversal): "...if one listens to temporal white noise and subsequently to the same signal after time reversal, it is not possible to distinguish not only which is which, but also whether they sound different."(González-Espinoza et al. (2020)).

IN THEORY, if the process (e.g. the future development of fuel consumption in a region, available electricity capacity of a transmission system) is not known, but various measurements, observations, and samples exist, one may accept neural networks as a suitable tool for energy forecasting or modelling. Rácz and Németh (2018)[10] applied neural networks this way, using a method called dynamic line rating when predicting the maximum transmission capacity of the electricity grid. Theoretical conditions were fulfilled; the process is only partially known; the standard models are based on a simplified heat equation neglecting, e.g. electromagnetic phenomena:[11]

[10] Rácz, L. and Németh, B. (2018). Investigation of dynamic electricity line rating based on neural networks. *Energetika*, 64(2)

[11] P_j - Joule heating (J)
P_s - solar heating (J)
P_c - convective cooling
P_r - radiative cooling

$$P_j + P_s = P_c + P_r \tag{3.1.1}$$

The above-mentioned measurements exist in the form of data for solar radiation, wind speed, and ambient temperatures of previous years and values of currents represented by the current and the temperature of the transmission line. The model examined above-ground lines with the maximum permissible temperature of 40°C due to the limitation of sag. Based on their results, the neural network calculated the temperature of the wire with an error of under 7%; this rate equals a maximum temperature difference of 2°C, which roughly coincides with the sensors' accuracy (Rácz and Németh (2018)). Answering the question of maximum transmission capacity for a new technology of installing lines *underground* would require a more complex model.

3.1.2 Energy modelling - examples

TO ILLUSTRATE THE MAIN IDEA, this work presents a typical modelling task in the energy sector: the contribution of energy system modelling to the transition towards renewable energy sources. Lund et al. (2017)[12] based this research question on two basic archetypes: simulation vs optimization of national energy models. While the optimization approach sets a pre-determined goal (e.g. the lowest energy consumption or the highest reduction of carbon dioxide emissions) and searches for the lowest-cost path possible to achieve it, simulation models create various scenarios for the future (figure 3.1). These scenarios take into account future technological options, fuel prices, and implemented policies. For example, they 1) test the resilience of technical infrastructure with hypothesized shocks, 2) show *a path* for achieving a low carbon energy future, or 3) test the hypothesis of countries' mitigation of gas import dependency on exporting countries through energy efficiency and renewable energy policies.

[12] Lund, H., Arler, F., Østergaard, P., Hvelplund, F., Connolly, D., Mathiesen, B., and Karnøe, P. (2017). Simulation versus optimisation: Theoretical positions in energy system modelling. *Energies*, 10(7):840. https://doi.org/10.3390/en10070840

As for forecasting, "the past matters" either equally as in simple moving average models or functions assign exponentially decreasing weights over time as in exponential smoothing models (figure 3.1).

Simply put, their role is definitely not to predict.

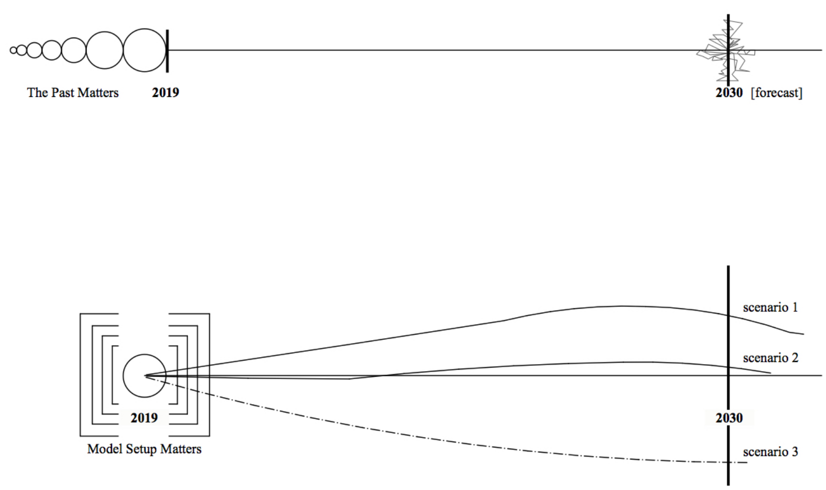


Figure 3.1: Energy forecasting vs energy modelling. Four layers in the model setup depict gas sector, energy sector of a country, economy of a country and the world economy.

To CONCLUDE, energy models cannot promise more than the representation of an energy sector in a macroeconomic model or of an entire national, regional, or global economy. Their outcome is a projection of various scenarios without probability distributions, as shown by Lund et al. (2017)'s exemplary research and in table 3.1. Strictly speaking, these cannot be labelled as forecasts. Fragkos et al. (2015)[13] attempted to solve this problem by integrating uncertainty into world energy modelling by using the PROMETHEUS Model. This hybrid model combines structural considerations (including expert judgement) with a time-series analysis to provide patterns of variation over time. Its deterministic version[14] has been extensively used in studies by the Eu-

[13] Fragkos, P., Kouvaritakis, N., and Capros, P. (2015). Incorporating uncertainty into world energy modelling: the PROMETHEUS model. *Environmental Modeling & Assessment*, 20(5):549–569

[14] In a deterministic version, all equation parameters are set at their mean versions.

ropean Commission, such as the EU Energy Roadmap 2050.[15]

[15] Shortly after, the district heating sector introduced its own Heat Roadmap of Europe because the EU Energy Roadmap 2050 anticipated the massive use of electricity in the heating sector.

3.1.3 *Analogy with music: An idea experiment*

At first glance, jazz piano seems like a spontaneous activity: A pianist is simply listening to jazz music in the background and a listener never knows what the final musical experience will be like. Likewise, those not familiar with computer science and statistics at a professional level (which even includes some energy experts and policy makers) may have the same impression when trying to understand forecasting methods. As a consequence, forecasts lack plausibility. The analogy with jazz improvisation works for the process of forecasting, but not for the outcome.

Upon closer inspection of jazz improvisation, one starts to question the aspect of *improvisation* in a performance - since there are certain rules (e.g. adapting to a rhythm, never playing the same note in a row, etc.) which guide a jazz pianist when improvising.

These rules are analogous to formulas in the classical Hubbert model, statistical methods and neural networks. An overview of these methods is given in figure 3.2. In free jazz[16], the more experienced the jazz pianist becomes, the fewer rules are applied, with the exception of one: feeling the music. The higher one climbs the ladder in figure 3.2, the more the music loses structure, while gaining mastery in becoming *unpredictable*. The same tendency is noticed in the development of forecasting methodology: from rather rigid methods to the soft computing trying to mimic intelligence. Thus, hidden layers in neural networks or transforming linguistically expressed knowledge (mirroring imprecise and uncertain human thinking) into workable algorithms (fuzzy logic) are placed higher up the ladder, although not achieving climax - a paradigm change to come. Single methods such as the Hubbert Curve Model or statistical models are being replaced by hybrid methods such as the Bayesian Model Averaging (Zhang and Yang (2015)) or the combination of the Autoregressive Moving Average (ARMA) method with genetic algorithms (Ervural et al. (2016)).

[16] In other jazz genres, most of improvisations follow more or less strict rules. Some rules are even related to the physique of a player and the used instrument.

The fact that most researchers[17] do not distinguish between *energy forecasting* and *energy modelling* is comparable to a situation when one listening to jazz music or a symphony refers to the activity as *listening* to the music. The reasons are as follows:

[17] As an exception, Dreborg (2004) in his dissertation distinguishes modelling and forecasting with his own terminology; forecasts are called "predictive approaches without a formal model of the system of interest."

In energy forecasting, models equal methods; in energy modelling - there is one energy model with the function of minimizing costs or maximizing welfare.

- The capability of making distinctions is related to intelligence and the understanding of the issue thanks *to having been exposed to* them

	Energy forecasting	**Energy Modelling**
Character	(mostly) stochastic	(mostly) deterministic
Examples of methods	a) econometric methods based on input-output coefficients b) added load methods considering additions and deductions to gas demand due to new connections and efficiency improvements	Optimisation based on maximising total surplus or minimising total costs for gas infrastructure
Statistics- exogenous data	Forecast load: weather variables (temperature and humidity), heating and cooling degree days, temperature-humidity index	Gas consumption (as a result of optimisation, it could be endogenous, too), price of gas and of alternative fuels for taking into account fuel-switch effects, supply constraints (e.g. pipeline capacities), existing technology stock.
Demand functions	(Mostly) weather responsive especially for short-term forecasting.	Price responsive, energy policies responsive
Human behaviour	Implicitly in the historical data	Researchers' own assumptions or assumptions taken from the official national models (e.g. energy scenarios from the UK Energy Research Centre's UK Energy 2050 project)
Geographical coverage	Irrelevant (mostly at city or state level) "Trading can take place anywhere with an internet connection."	a) global e.g. Shell global energy model from 1974, World gas model; PROMETHEUS b) regional, national c) supranational, e.g. European Gas Market Model
Outcome	Prediction (point, interval, density)	a) energy projection as a function of energy policies applied b) projection scenario (set of assumptions on key inputs) without a probability distribution
Evaluation of the outcome	Root Mean Squared Error (RMSE) Mean Absolute Percentage Error (MAPE) Weighted Mean Absolute Percentage Error (WMAPE)	No general evaluation standard
Client/ contracting body	Private Sector. Academia.	Public sector: federal agencies, ministries, academia.

Table 3.1: Energy forecasting vs. energy modelling with a focus on the gas sector

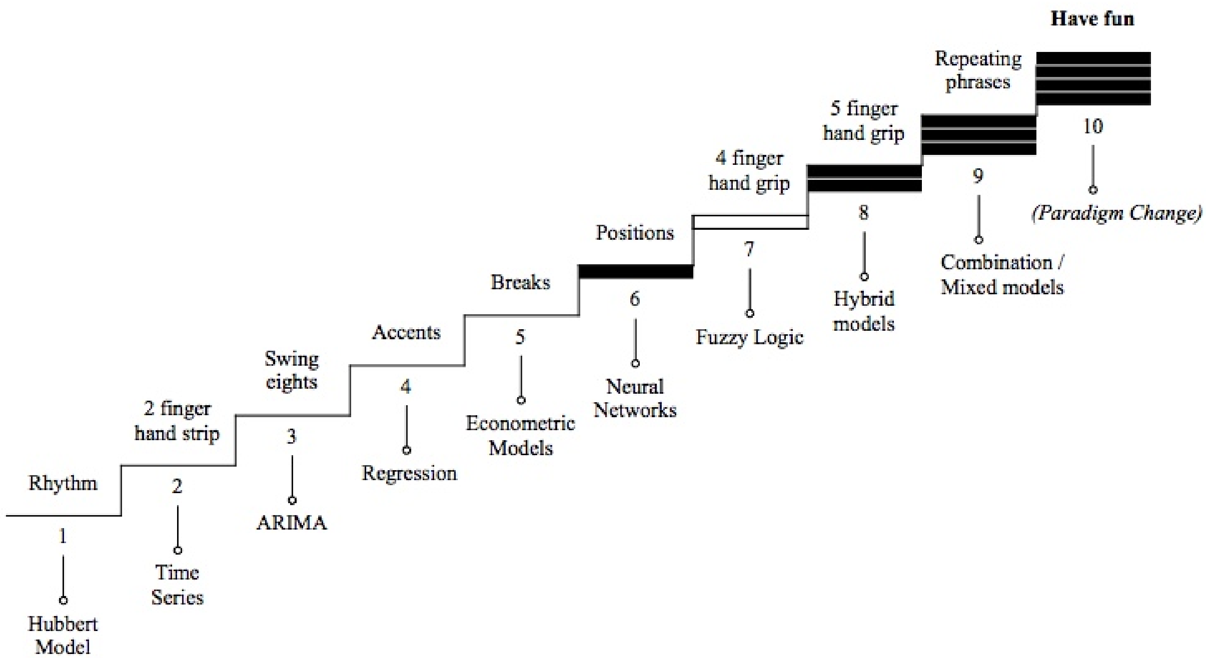


Figure 3.2: Parallels between steps for learning jazz piano improvisation and developing new methods for energy forecasting. Sources: Maria Greitzer (analogy), Oliver Prehn (10 Steps in Piano Jazz Improvisation Prehn (2018)), Comfort Mosha (Illustration).

before.

- Sharp thinking in mathematical terms cannot be applied to the language used.

3.1.4 *Interdisciplinary Aspects*

STATISTICAL LEARNING MODELS DOMINATE the current practices in energy forecasting and modelling. The influence of the social sciences on forecasting practices is illustrated in an analogy to a piano arrangement in figure 3.3.[18]

[18] Cage, J. (1952). 4′33 Piano solo arrangement. First movement. https://musescore.com/user/5832946/scores/1559096 Last accessed 2020-07-26

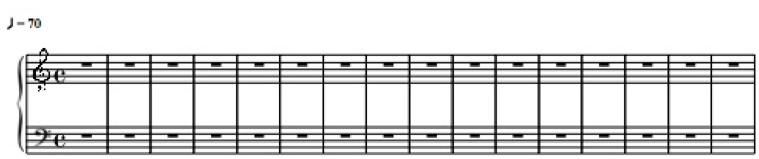


Figure 3.3: Illustration of the role of social science in energy forecasting. Source: Music note for 4´33 Piano solo arrangement, First movement. Composed by John Cage.

Because of the consequences of the status quo shown in figure 3.3, Sovacool et al. (2015)[19] suggested a program-centred approach to the energy field instead of the technology-centred approach to encourage interdisciplinary depth. Although this direction is already noticeable in the calls for research projects of national institutions in Germany, energy forecasting and modelling has not been reached by this trend yet. What is needed is shifting the attention from questions such as "*how to demonstrate that neural networks are the right tool for prediction of gas demand in Germany in* 2030" to "*how high will the gas peak demand be in Germany in* 2030?"

[19] Sovacool, B. K., Ryan, S. E., Stern, P. C., Janda, K., Rochlin, G., Spreng, D., Pasqualetti, M. J., Wilhite, H., and Lutzenhiser, L. (2015). Integrating social science in energy research. *Energy Research & Social Science*, 6:95–99

Regarding interdisciplinary aspects, International Energy Relations discuss energy supply security and the reliability of energy models. Jefferson (2016)[20] calls the discipline *International Political Economy of Energy*[21] having its roots in the 1970*s* as a product of the OPEC crisis. Gas crises in Europe in 2009 and 2014 had fewer effects on energy policies in Europe, but initiated interest in energy models' setup for testing the availability of gas in Europe or the level of dependency on imports to name few. Studies were mostly conducted for vulnerable countries in Central and South-Eastern Europe and for the United Kingdom.

[20] Jefferson, M. (2016). Energy realities or modelling: Which is more useful in a world of internal contradictions? *Energy Research & Social Science*, 22:1–6. `https://doi.org/10.1016/j.erss.2016.08.006`

[21] Articles of this kind are to be found in Journals such as Energy Research and Social Science, or Energy Policy.

Interdisciplinary interdependencies influence the outcome of energy modelling as the following example shows. China's gas reserves are overestimated due to differences in Chinese and internationally accepted definitions of gas resources (i.e. gas estimates in the ground) and reserves (i.e. gas produced with current prices and technology) (Shaikh and Ji (2016)).[22] Changed data of the BP, the US Department of Energy or the IEA serve as inputs in energy models and stimulate discussions of *Peak Demand* instead of Peak Oil. Some global energy models already included the concept of Peak Demand, e.g. DNV GL already forecasted an oil demand peak for 2023 and gas demand peak for 2035 in their Energy Transition Outlook in 2018 DNV-GL (2019).[23] DNV GL characterizes their model as system dynamic modelling of the world energy system.[24] Continuation of current technology trends is assumed with one exception being the increased use of hydrogen for energy purposes. Other global energy models follow expectations regarding the use of hydrogen, too. For example Prometheus (in Fragkos et al. (2015))[25] identifies 18 hydrogen production technologies in a separate hydrogen module.

[22] Shaikh, F. and Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, 77:25–32

[23] DNV-GL (2019). Energy Transition Outlook. Oil and Gas Forecast 2050. `https://eto.dnv.com/2017/oilgas` Last accessed 2021-03-26

[24] System dynamics is a branch of systems theory (a way to see the whole as the collection of its interacting parts) "that recognises the role of positive and negative feedback, in which systems can spin out of control, as in virtuous or vicious cycles, and in which systems can be kept within bounds, respectively." (Bale et al. (2015)). Wang et al. (2019) considers DNV GL the only one major energy institute that does not conduct a demand-driven analysis, meaning assuming that gas resources will automatically meet the future demand.

[25] Fragkos, P., Kouvaritakis, N., and Capros, P. (2015). Incorporating uncertainty into world energy modelling: the PROMETHEUS model. *Environmental Modeling & Assessment*, 20(5):549–569

For the latest list of *global gas demand* scenarios from IEA, BP, ExxonMobil, Equinor, DNV GL, EIA, and Shell, see Bradshaw and Boersma (2020).

3.1.5 Conclusion

Most papers follow the same structure:

1. A short reasoning for choosing gas forecasting as the research object, e.g. the rapid growth of gas consumption in a country, or a relevance of the precise forecasting for economic progress;
2. A description of the research methodology used as well as alternative methods;
3. Proof that the chosen methodology outperforms other methods in terms of MAPE and RMSE, or by naming novel qualities of a suggested method (e.g. dynamic, adaptive).

There are some peculiarities of forecasting gas demand when compared to oil or electricity demand forecasting. Short-term gas load (consumption) is divided into heat load, dependent mostly on outdoor temperature and base load.[26] The relation between heat load and temperature is linear within a certain range of temperatures as shown e.g. in Merkel et al. (2018) for several Midwestern US operating areas or in Franco and Fantozzi (2015) for temperatures below 15°C in Italy. Most forecasters treat the whole gas demand as a temperature-dependent output.

[26] The base load covers other usages of gas without direct dependence on the temperature, especially in the industry sector.

Research on gas modelling contributes rather to the methodology development[27] than to the further knowledge gain for the gas sector. If planners followed forecasts from 2015, China for example would be dealing with a remarkable oversupply of gas right now. However, it is the underestimation of gas consumption that can threaten the economy and the welfare of the population.

[27] Analysts do recognize shortcomings of their methods when the values violate economic theory (i.e. logic of the models) or common sense. This is usually solved by arbitrary interventions or simply by ignoring these values.

As for energy security, the impact of research projects is hardly measurable as policy makers form their policies in line with studies assigned by ministries. Research projects become secondary literature for them.

3.2 Related work

This section reviews literature on gas forecasting with an emphasis on the essential idea of forecasting methods and accuracy criteria. The criteria measure the out-of-sample (or test) error, i.e. the prediction error over an independent test sample. Formally, y denotes the output (a target variable), X the vector of inputs and $\hat{f}(x)$ the prediction model from a training data set τ. The loss function is denoted by $L(y, \hat{f}(X))$ (Hastie et al. (2009)). Typical loss functions work with absolute or squared values. For the test error Err_τ[28] for a fixed training set τ we have

[28] Training error and its relation to complexity is described in the section on complexity.

$$Err_\tau = E[L(y, \hat{f}(X))|\tau] \tag{3.2.1}$$

In simpler terms, a forecast error is defined as

$$Err = y - \hat{f}(X) \tag{3.2.2}$$

3.2.1 Forecasting assumptions

Lindley (2001)[29] stated that:
"a model is merely your reflection of reality and, like probability, it describes neither you nor the world, but only a relationship between you and that world." This work adds *a simplified reflection*. All forecasting methods are based on the assumption of underlying relationships among multiple variables, and their discovery assists the estimation of the response variable[30] in the future. Moreover, it is assumed that data quality is sufficient and representative of a research object. Additionally, if the model[31] predicts accurately in one forecasting period, it will also produce accurate forecasts in *further periods*. Besides these implicit assumptions, a few more premises set the trend in forecasting and change almost every two decades. In the second half of the 20^{th} century, statisticians believed there could be one superior model that fit almost all forecasting problems. Supposedly, it was just a matter of time before it was found.

[29] Lindley, D. V. (2001). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337. https://doi.org/10.1111/1467-9884.00238

[30] Response variables (statistics) and outputs (machine learning) are considered as synonyms.

[31] A model can be understood as *"a smooth, low-order polynomial curve fitted to a cloud of points on a plane"* as in Zellner et al. (2002).

Zellner, A., Keuzenkamp, H. A., and McAleer, M. (2002). *Simplicity, inference and modeling: Keeping it sophisticatedly simple*. Cambridge University Press, Cambridge and New York. ISBN 0521803616

Regarding complexity, data scientists take clashing positions; either they assume that "*more sophisticated models lead to increases in the prediction accuracy*" or "*simple models predict at least as accurate as complex models and in most cases, they outperform complex models*". The definition of a model sophistication has not been defined in the *classical* literature on forecasting; non-parametric non-linear models are perceived to be complex. Kaposty et al. (2020) describes more complex methods as models able to take more information into consideration.

For parametric models, there is an assumption of uncertainty following a given probability distribution, this is not the case for non-parametric models.

3.2.2 *Accuracy criteria*

DEPENDING ON THE TASK, the accuracy is computed by comparing forecasts to their real values with the aid of commonly used forecast measures as described in the subsection below or - for model selection - to a benchmark model that is usually a persistence model or the mean of previous values. Let r_t denote the relative error, then

In the persistence model, forecast equals the latest observation in time series.

$$r_t = e_t / e_t* \tag{3.2.3}$$

where e_t means the forecast error of a model compared and e_t* the forecast error of the benchmark method (Hyndman and Koehler (2006)).

ALTHOUGH OVERLOOKED IN THE LITERATURE, Kuhn and Johnson (2016)[32] in their textbook *Applied Predictive Modelling* discuss the upper bound of the accuracy related to the response variable. Systematic error is an inevitable part of any error and includes measurements, summing and reporting final values. As for a self-constructed data set for forecasting gas imports to Germany, imperfections include measurements of amount of gas transported at import nodes, reporting issues and changes in the methodology of calculations.

[32] Kuhn, M. and Johnson, K. (2016). *Applied predictive modeling*. Springer, New York, Corrected 5th printing edition. ISBN 978-1-4614-6849-3

Forecasting criteria

The forecasting criteria below are mostly used for measuring accuracy of univariate time series forecasts, i.e. forecasts for time series that consists of single scalar observations measured sequentially over equal time steps.

Automatic forecasting software may be set to exclude models with results exceeding 50 per cent in MAPE (Katz (2020)).

a) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) is a scale-independent measure used for the comparison of models for forecasting variables with different mean value; with the range of 1-10 (highly accurate), 10-30 (good forecast) and above 50 (inaccurate forecast).

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right| \tag{3.2.4}$$

where n is the number of data points, y_t the actual value and e_t the difference between an actual value y_t and the predicted value $\hat{y}_t$. Since this measure penalises positive errors more than negative ones; Makridakis (1993)[33] suggested a *symmetric* mean absolute percentage error:

[33] Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4):527–529. https://doi.org/10.1016/0169-2070(93)90079-3

$$sMAPE = mean(200|y_t - \hat{y}_t| / (y_t + \hat{y}_t)) \tag{3.2.5}$$

sMAPE criterion can take negative values and it is also not fully *symmetric* (Hyndman and Koehler (2006)).

b) ROOT MEAN SQUARE ERROR (RMSE) is used in linear models, e.g. for electricity production from wind power plants based on meteorological models, as its minimization (a quadratic problem) can be solved by differentiation or iteratively by gradient descent (Browell (2015)). Aggregation and the mean calculation mask the behaviour of various models; scatter plots display the insensitivity of the RMSE (as an aggregate number) as shown in Rieck (2017).[34] Besides practice in forecasting, the RMSE is also used as a quality measure for dimensionality reduction methods.

Gradient descent is a generic approach to minimizing in-sample error $R(\theta)$ in neural networks. Neural networks with no hidden layer are actually linear multinomial regression models (Hastie et al. (2009)).

[34] Rieck, B. A. (2017). *Persistent Homology in Multivariate Data Visualization*. PhD thesis, Heidelberg University Library. http://archiv.ub.uni-heidelberg.de/volltextserver/22914/1/Dissertation.pdf Last accessed 2021-04-13

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{3.2.6}$$

c) Mean absolute error (MAE) measures the mean absolute difference between the predicted and observed values. The error is representative in situations where "*the economic cost of a forecast error is proportional to the magnitude of the error, as opposed to its square*" (Browell (2015)).[35] However, for this case, no real example from the energy sector was found. The mean absolute error is also used for expressing the forecast error in relation to other errors.

[35] Browell, J. (2015). *Spatio-temporal prediction of wind fields*. PhD thesis, University of Strathclyde. http://oleg.lib.strath.ac.uk:80/R/?func=dbin-jump-full&object_id=25822 Last accessed 2020-12-20

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{3.2.7}$$

d) Mean square error (MSE), as every error, could be decomposed into a bias term and variance. The bias is the consistent offset of the forecast, the variance represents the variation of the forecast *error* around its mean.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{3.2.8}$$

Both, MAE and MSE are generalized by the Minkowski objective function with the exponent R (Pelikán (2014)).[36] For $R = 1$ the MAE is computed, for $R = 2$ it is the MSE. All three errors (RMSE, MAE and MSE) are scale-dependent and sensitive to outliers.

[36] Pelikán, E. (2014). Forecasting of processes in complex systems for real-world problems. Tutorial. *Neural Network World*, 24(6):567–589. http://www.nnw.cz/doi/2014/NNW.2014.24.032.pdf

$$L_{\mathrm{M}} = \frac{1}{n}\sum_{t=1}^{n} e_t^R \tag{3.2.9}$$

e) The coefficient of determination R^2 measures the proportion of the variance that can be explained by the model. There are many ways to compute R^2: one of them, correlation coefficient, measures how well the predicted and real (measured) values are correlated. The most used form of the coefficient is Pearson's correlation coefficient r.

$$r = \frac{n \sum XY - (\sum X \sum Y)}{\sqrt{[n \sum x^2 - (\sum x^2)][n \sum y^2 - (\sum y^2)]}} \tag{3.2.10}$$

where n denotes number of observations in the regression equation, X the mean of the independent variable of the regression equation and Y the mean of the response variable (output). As follows from the formula, the correlation coefficient does not provide any information on systematic over- and underpredictions of a model.

f) "Hit ratio" reflects the proportion of successful forecasts, defined by the match of their algebraic sign (e.g. a percentage change of gas price compared to its previous value) to the true value. It is rare to find the measure *hit ratio* in energy forecasting; for example Busse et al. (2012)[37] used it for measuring accuracy in gas spot price forecasting.

$$hitratio = \frac{forecasts - errorcount}{forecasts} \tag{3.2.11}$$

Alternative accuracy measures include the MASE, the scaled variant of the mean absolute error (MAE), proposed by Hyndman and Kochler, where the scaling is equal to the MAE of the seasonal random walk for the in-sample data[38] or the overall weighted average (OWA) of chosen accuracy measures. Here, outliers[39] are less penalized than in the case of RMSE where the deviation from the actual value is squared.

3.2.3 *The essential idea of methods*

Studying forecasting literature of the last 20 years, it is hard to ignore the number of studies devoted to China or Turkey. It is speculated that this is due to their rapidly increasing gas demand due to population increase (Turkey) or high economic expectations in the last decade (China). The paradox lies in the realization that most forecasting methods work best under stable *business as usual* conditions.

Second, the setup of every study with its forecasting models is unique in almost all criteria:

a) Training and test data sets in terms of their size and quality.

b) Goal of forecasting and various time horizons. Authors of reviews choose arbitrary division lines; Ghalehkhondabi et al. (2017)[40] understand short-term forecasting as forecasting up to a month, mid-term forecasting up to five years and long-term forecasting from five

[37] Busse, S., Helmholz, P., and Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - an analysis of potential influence factors on basis of a NARX neural network. *Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012*. https://publikationsserver.tu-braunschweig.de/servlets/MCRFileNodeServlet/dbbs_derivate_00027726/Beitrag299.pdf Last accessed 2021-04-14

[38] Petropoulos, F., Hyndman, R. J., and Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554

[39] Data points (cases) with a long distance to all other cases.

[40] Ghalehkhondabi, I., Ardjmand, E., Weckman, G. R., and Young, W. A. (2017). An overview of energy demand forecasting methods published in 2005–2015. *Energy Systems*, 8:411–447. 10.1007/s12667-016-0203-y

to 20 years. Debnath and Mourshed (2018) define a short-term horizon as up to three years, medium-term from three to fifteen years and the long-term as over fifteen years. The short-term forecasts in the energy sector are the matter of interest for energy traders[41] and distributors of electricity; therefore, in the *electricity load* forecasting, the short-term period covers less than one hour and long-term more than one-year.

[41] For example, day-ahead electricity price forecasts in Gianfreda et al. (2020).

ABOVE 20 YEARS, either a long-term forecast estimate e.g. the return on investment into the technology (new gas turbines, wind power turbines) or the whole energy sector, economy or global economy are modelled by applying the function of the minimal cost of produced and/or distributed energy, published by the IEA, the US EIA and others. Since governments partly control the outcome of demand of energy carriers, the IEA and other organizations publish projections under Scenarios. Geographical coverage affects the choice of a forecasting horizon and required accuracy: the higher the level (district, city, country, global), the higher the tendency to forecast long-term and the lower accuracy is expected given the uncertainty of the input information.

"Reference Scenario" - the continuation of already applied policies, "New Policy Scenario" - reflecting the change in policies.

IN CASE STUDIES, the length of forecast horizons depends on the characteristics of the object's field. Table 3.2 contains forecast horizons, common temporal resolutions and decisions relevant for wind energy from the dissertation of Browell (2015).[42] Two more columns have been added to show if and how forecasts could be applicable for energy networks: district heating (DH) network with the spatial boundary of a city and a gas transmission network with nodes at the countries' borders. Whereas the wind power plant operators balance their electricity supply in the very-short-term, district heating operators can have up to a few hours to react to changes in demand due to system latency. In large networks at city level, there is a time lag of up to 12 hours for delivering heat from a supplier to the final consumer. Predicting hourly heat demand depends more on behavioral patterns than on buildings' thermal properties. Similar to district heating networks, gas networks contain *linepack*, the amount of energy in pipelines, computed with the fixed temperature and the average heating value of gas (Botev and Johnson (2020)).[43]

[42] Browell, J. (2015). *Spatio-temporal prediction of wind fields*. PhD thesis, University of Strathclyde. `http://oleg.lib.strath.ac.uk:80/R/?func=dbin-jump-full&object_id=25822` Last accessed 2020-12-20

[43] Botev, L. and Johnson, P. (2020). Applications of statistical process control in the management of unaccounted for gas. *Journal of Natural Gas Science and Engineering*, 76:103194

c) The choice of variables

A variable neglected by one author is of importance in other studies, which is understandable if this work draws a line between in-

	Forecast horizon	Resolution	Decision	District heating (DH) networks Prediction: Heat demand for residential sector, based on weather	Gas networks Prediction: gas demand
Ultra-short-term	<1 min,	Seconds	Wind turbine control	Not applicable	Not applicable
Very-short-term	<1 hour	1,5,10,15 min.	Balancing, wind farm control, spot markets	Not applicable for networks at the city level; a DH network is a short-term storage.	
Short-term	1-48 hours	30 min., 1 hour	Generation scheduling, day-ahead markets, some spot markets	Balancing the supply	Balancing the supply
Medium-term	1-10 days	1 hour, 3 hours	Generation scheduling, maintenance planning		Balancing the supply
Long-term	Months-years	Days-months	Maintenance planning, resource assessment/project financing	Maintenance planning, fuel purchase	Maintenance planning, Strategic planning, Long-term international contracts, LNG contracts.

Table 3.2: Forecast horizons, common temporal resolutions, decisions for wind energy (Browell (2015)), district heat and gas networks (M.Greitzer)

dustrial and post-industrial countries. The increase of average income/per capita in China could cause increased gas consumption of the country, whereas in Germany, this linear relationship does not exist.

In the next section, models are discussed at length.

Models, methods and techniques are treated as synonyms.

THERE ARE NO UNIFIED CRITERIA for models' categorization. For comprehensive reviews on models in gas forecasting, see Soldo (2012), Tamba et al. (2018) and Sen et al. (2019). Šebalj, Dario and Dujak Davor and Mesaric Josip (2017)[44] reviewed 187 papers on gas *consumption* forecasting and listed the nine most used methods:[45] neural networks, adaptive neuro-fuzzy inference system (ANFIS), grey model, support vector machine (SVM)[46], genetic algorithms[47], mathematical and statistical models, time series, and hybrid models.[48] Szoplik (2015)[49] provided an overview of three forecasting methods with a practical meaning for predicting gas *consumption*: time series methods, regression models, and neural networks. These reviews show that most forecasts in the energy field deal with energy prices (Herrera et al. (2019)), fuel production (Semenychev et al. (2014)) or fuel consumption of natural gas and coal, as both belong to finite energy resources.

Tamba et al. (2018) reviewed models up to 2015 and ordered them chronologically from oldest to most recent:

- Hubbert model
- statistical models (ARIMA, time series models, decomposition models and trend analysis)
- regression models
- econometric models
- AI-expert systems (neural networks)
- fuzzy logic
- grey prediction models
- genetic algorithms
- mathematical models
- hybrid models
- combination or mixed models.

Sen et al. (2019) grouped forecasting methodologies by the period of their use into three types: ARIMA modelling, decomposition approaches on a daily basis and heuristic approaches based on economic indicators such as GDP, population, and inflation.

[44] Šebalj, Dario and Dujak Davor and Mesaric Josip (2017). Predicting natural gas consumption - a literature review. http://archive.ceciis.foi.hr/app/public/conferences/2017/08/SPDM-2.pdf

[45] Side notes refer to the newest studies of models for gas forecasting.

[46] Wang, Z., Li, Y., Feng, Z., and Wen, K. (2019). Natural gas consumption forecasting model based on coal-to-gas project in China. *Global Energy Interconnection*, 2(5):429–435

[47] Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. *Energy*, 178:585–597

[48] Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. *Energy*, 178:585–597

[49] Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208–220

"The object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data." in:

Fischer, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368

[50] Erdoğdu, E. (2010). Natural gas demand in Turkey. *Applied Energy*, 87(1):211–219. https://doi.org/10.1016/j.apenergy.2009.07.006

For time series data, Erdoğdu (2010)[50] listed the following approaches to economic forecasting: a) exponential smoothing methods, b) single-equation regression models, c) simultaneous-equation regression models, d) ARIMA, and e) vector autoregression.

Within the forecasting community, various criteria are used to make a distinction between *statistical models* and *machine learning models*. In the broad sense, statistical methods are described as "variants of exponential smoothing and ARIMA methods" in Makridakis et al. (2018)[51] and machine learning methods are anything else, covering neural networks and random forests. A statistical model can also be understood as a model learning its parameters in one series at a time, whereas the machine learning (ML) model finds the parameters across multiple series. Barker (2020)[52] suggests *a new division* between structured models (such as an autoregressive model) and unstructured models (neural networks) based on the way the data are generated: is the process of generation defined a priori or is it learned from the data?

[51] Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808

[52] Barker, J. (2020). Machine learning in m4: What makes a good unstructured model? *International Journal of Forecasting*, 36(1):150–155

Januschowski et al. (2020)[53] formulated objective and subjective dimensions of the classification; the next paragraph shortly summarizes both groups according to their paper, *with comments*.

[53] Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., and Callot, L. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36(1):167–177

3.2.4 Objective methods according to Januschowski et al. (2020)

Global and local methods. Local methods mean estimating parameters for a model in each time series separately; global methods search across time series.

Point vs probabilistic forecasts.[54] Point forecasts assign one single value for the amount category of a research object (e.g. wind speed, gas consumption) for a forecasting horizon. Their main advantage is how quickly the information is grasped by the public, though this is without any quantification of forecast uncertainty. Probabilistic forecasts estimate the likelihood of possible outcomes and thus, they are combinations of the two qualities: sharpness and reliability.

[54] Interval and quantile forecasts

In gas forecasting, researchers' results take the form of point forecasts as these are subjects of their contracts with contractors.[55] *However, interval forecasts would provide higher information value as it is the maximum demand that is relevant for maintaining the security of energy supply. Second, overprediction and underprediction of gas consumption do not have the same negative impact in terms of economic efficiency and security. Therefore, quantile forecasts are more appropriate since they estimate the probability that an observation will exceed the set value.*

[55] In German Auftraggeber, usually Ministries and other public institutions.

A step further denotes probabilistic forecasting. Density forecasts estimate probability distribution of the possible future values of a variable meaning the forecast contains information regarding the full range of possible outcomes (Browell (2015)).

COMPUTATIONAL COMPLEXITY, decreased by parallelizability. *Complexity measured by the time spent on producing the forecast serves as a main criterion for assessing complexity in forecasting. Still, it only covers a single aspect of the complexity phenomenon.*

LINEARITY AND CONVEXITY. Labelling statistical models as linear and machine learning models as non-linear over-simplifies the categorization. Methods applying a convex loss[56] function for an optimization procedure can be easily explained logically and thus better embedded in a theory. Non-convex models may produce impressive results, however there is no logical explanation as to how they work.

[56] A loss function $L(y, \hat{y})$ defines the estimated prediction error by measuring how close $\hat{y}$ is to y.

Subjective dimensions according to Januschowski et al. (2020)

DATA-DRIVEN VS MODEL-DRIVEN METHODS. Machine learning (ML) models are perceived as data-driven models if the pattern comes from the data only, with the risk of over-fitting. Additional information in ML models take the form of adding a) external variables, b) features from descriptive statistics (spectral entropy) or c) features automatically generated by other ML algorithms. Data pre-processing affects model performance.

ARIMA models are typical representatives of model-driven methods with defined assumptions.

In gas forecasting research studies, data-driven models start to prevail while model-driven models are used as benchmarks.

ENSEMBLE VS SINGLE MODELS

DISCRIMINATIVE VS GENERATIVE

STATISTICAL GUARANTEES ARE EXPLAINED in articles published in statistical and ML journals. In statistical journals, theoretical assumptions of a new proposed method are supported by an empirical study based on a small-scale data set to validate theoretical results. In the next step, results are expected to outperform results from previous studies. In ML journals, a new procedure is described as well as the motivation to use it; the proposal is contextualized within the existing literature (if possible) and then the method is tested on several standard data sets.

In energy journals, first the motivation for forecasting gas demand is stated by two arguments – if gas demand is predicted at state level, it is the increased security of energy supply and capacity building (pipelines, storage); in case of prediction of gas demand for a part of a distribution network, energy and financial efficiency comes into play. Second, the ultimate goal of forecasting is not to pursue the development of the forecasting science but to generate predictions for demand. Thus, forecasting models are perceived only as a tool, one out of many, including for example expert judgements derived from 10-year development plans for gas infrastructure. Researchers follow recent trends in forecasting and apply machine learning models but stay interested in the interpretation of the forecast (and not the forecasting model itself). Recalling the first two sentences in the introduction, there is often no evidence of utilizing new forecasts in policy making.

EXPLANATORY/INTERPRETABLE VS PREDICTIVE. Machine learning methods are often seen as black boxes. Wiener (2013)[57] understood them as “a piece of apparatus, such as four-terminal networks with two input and two output terminals, which performs a definite operation on the present and past of the input potential, but for which we do not necessarily have any information of the structure by which this operation is performed". He defined the white box in the same fashion with an essential distinction: *"we (researchers) are able to follow the determined process of gaining outputs as we built in the relation between input and output."*

[57] Wiener, N. (2013). *Cybernetics or control and communication in the animal and the machine*. MIT Press, Cambridge, Mass., 2. ed. edition. ISBN 9781614275022

Within the gas sector, results coming from black boxes are not acceptable. Models from the IEA, the BP and the U.S. EIA produce explanatory forecasts (predictions with scenarios) by modelling the whole gas sector as a part of the modelled economy and following the goal of minimized costs.

The International Energy Agency
Energy Information Administration

OUTSIDE THE FORECASTING COMMUNITY, figure 3.4 shows a visualization of methods used in studies of energy and water demand, presented in Sharmina et al. (2019).[58] Based on three references, Sharmina et al. (2019) state the machine learning methods are more accurate than statistical methods. The reason the latter prevails regardless is because of the built-in agenda of a technical discipline, where authors borrow the term "shared technical interest" from Asdal (2011).

[58] Sharmina, M., Abi Ghanem, D., Browne, A. L., Hall, S. M., Mylan, J., Petrova, S., and Wood, R. (2019). Envisioning surprises: How social sciences could help models represent ‘deep uncertainty’ in future energy and water demand. *Energy Research & Social Science*, 50:18–28. https://doi.org/10.1016/j.erss.2018.11.008

The next section describes chosen models and their applications in gas forecasting.

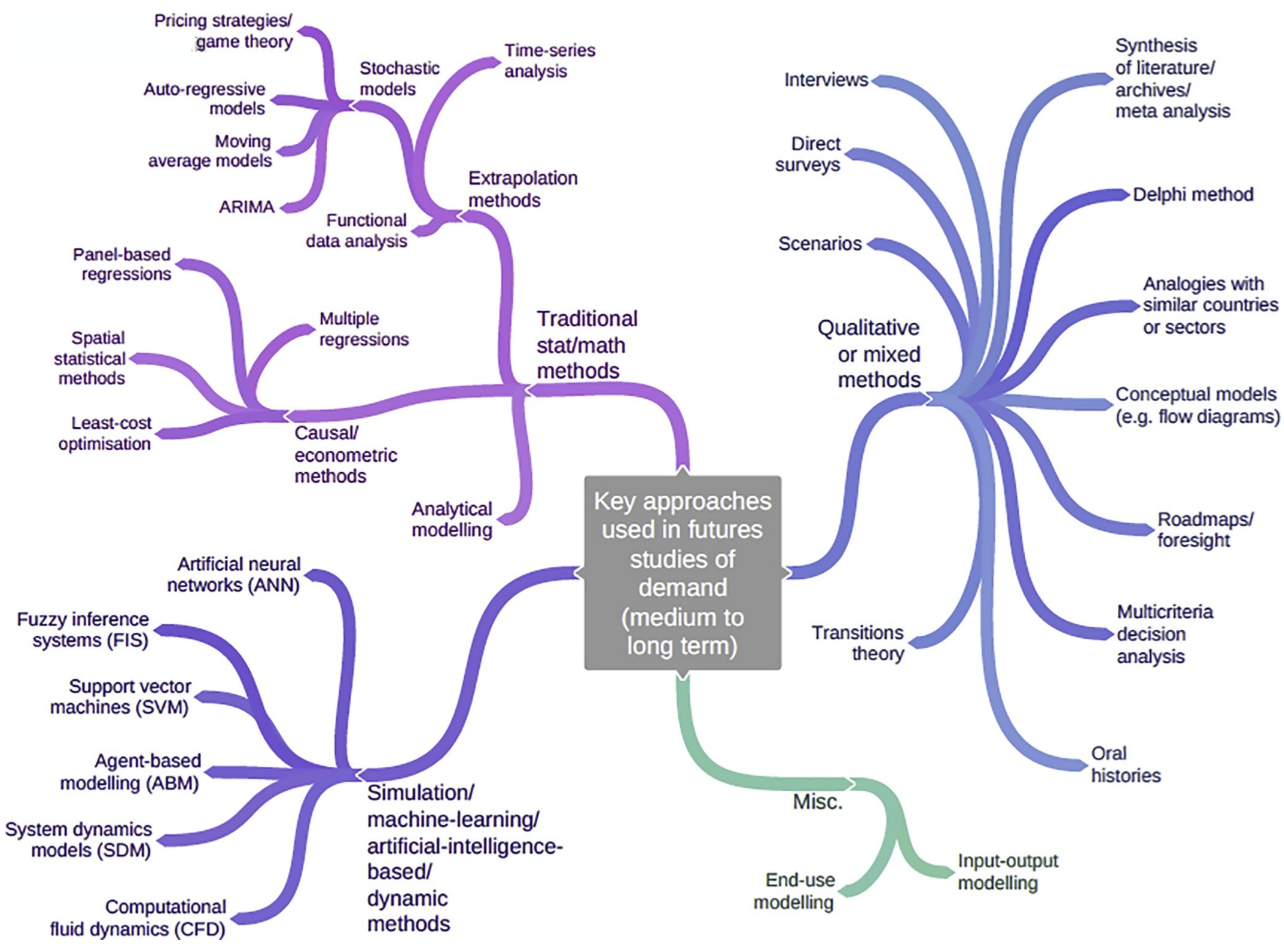


Figure 3.4: Visualisation of methods used for energy and water demand.

3.2.5 Forecasting models and applications

Hubbert Model – a theoretical production model with various adaptations for yearly forecasts of *gas production*. The classical version of the model assumes a symmetric production curve, i.e. the production peaks at a point when half of the ultimately recoverable resources are already exploited (Wang et al. (2019)).[59] Real production curves show a higher rate of production before the peak is reached than after the climax point (Wang et al. (2019)).

In gas forecasting, Siemek et al. (2003)[60] used the Startzman modification of the Hubbert model[61] to describe possible gas demand in Poland, based on economic development and taking into account production/demand maxima of energy carriers. For forecasting gas production in China, Ma and Li (2010)[62] excluded this model from their list of methods reasoning that "the predicted outputs are generally greater than actual outputs and the prediction results are less accu-

[59] Wang, Z., Li, Y., Feng, Z., and Wen, K. (2019). Natural gas consumption forecasting model based on coal-to-gas project in China. *Global Energy Interconnection*, 2(5):429–435

[60] Siemek, J., Nagy, S., and Rychlicki, S. (2003). Estimation of natural gas consumption in Poland based on the logistic-curve interpretation. *Applied Energy*, 75(1-2):1–7

[61] Al-Fattah and Startzman (1999) modified the Hubbert model assuming *multiple* gas production cycles.

[62] Ma, Y. and Li, Y. (2010). Analysis of the supply-demand status of China's natural gas to 2020. *Petroleum Science*, 7(1):132–135. https://link.springer.com/content/pdf/10.1007/s12182-010-0017-9.pdf

rate with time." This loss of accuracy is the nature of every forecasting method.

IN CONTRAST TO THE ELECTRICITY SECTOR, global natural gas production does not have to match gas demand *in the same period* due to the option to store gas[63] and other alternative fuels (synthetic methane, hydrogen). Still, similarities between the electricity and gas distribution systems (grid-bounded, dependence on temperature)[64] serve as justification for applying models from electricity (load) forecasting to predicting gas spot prices as argued by Busse et al. (2012).

[63] To calculate the total storage capacity, cushion gas is subtracted from the amount of stored gas. Cushion gas is the amount of gas required for maintaining adequate underground storage reservoir pressures and deliverability rates throughout the output cycle.

[64] Dependence on temperature is country-specific (more details in the modelling section on forecasting gas demand in Germany).

1. THE PERSISTENCE MODEL, also called a naïve forecast model, is the simplest reference model supposing that the response variable (output) y at some point in the future ($t + \Delta$) will be the same as the observation x at time t.

$$y + \Delta = x \tag{3.2.12}$$

   Browell (2015)[65] writes "The performance of the persistence forecast is considered a measure of the 'predictability' of a particular times series and is still used by some practitioners in the energy industry today for short-time forecasting." If any model does not outperform the persistence forecast, additional effort and cost in terms of computational time is not justified. New models do not have to outperform a naïve forecast remarkably; e.g. in Busse et al. (2012) recurrent neural networks[66] achieved the best performance with the hit ratio accuracy measure of 0.64 and the MAPE of 88.20 whereas the persistent forecast would deliver a hit ratio of 0.6 and the MAPE of 317.

[65] Browell, J. (2015). *Spatio-temporal prediction of wind fields*. PhD thesis, University of Strathclyde. http://oleg.lib.strath.ac.uk:80/R/?func=dbin-jump-full&object_id=25822 Last accessed 2020-12-20

[66] In recurrent neural networks, the gradient may vanish during backpropagation. The Long Short-term memory (LSTM) algorithm (deep learning) fixes the short memory of recurrent neural networks (Anagnostis et al. (2019)).

2. STATISTICAL/TIME SERIES MODELS. Time series models imply that the history, mirrored in data sets, has its own pattern that will probably continue into the future.

   IN JUDGEMENTAL FORECASTING, the very short length of time series (an observation) or relatively long observations (more than 20 observations) will provide better results than, for example, the last five observations (Theocharis and Harvey (2019)).[67]

   Time-series analysis assumes:

   (a) Random error terms are normally distributed. Random error is an estimate of the variance of an output; this variance emerges due to factors not included in the model.

[67] Theocharis, Z. and Harvey, N. (2019). When does more mean worse? Accuracy of judgmental forecasting is nonlinearly related to length of data series. *Omega*, 87:10–19

(b) There are dependable correlations between the variable to forecast and other independent variables.

(c) Past patterns in the variable to be predicted will continue unchanged into the future.[68]

(d) The data does not exhibit a trend.

[68] Policies such as coal-to-gas program in Northern China (see Wang et al. (2019)) invalidate this assumption.

3. STATISTICAL/AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODELS $(p, d, q)(P, D, Q)_S$[69]

These models are typical representatives of model-driven methods with defined assumptions; many *automatic* forecasting algorithms are based on them. An ARIMA model seeks additive relationships between lags of a series and its future values[70] under the condition of stationary time series (identically distributed). The methodology of a model consists of four steps:

- identification of values of a model by tools such as an autocorrelation function (ACF), or a partial autocorrelation function (PACF);
- estimation;
- checking residuals from the estimation to identify any autocorrelations and partial correlations of the residuals that are statistically significant;
- forecasting.

[69] In notation, p stands for the order of the auto-regression, d is the order of the differencing, and q stands for the moving average process. Capital letters are seasonal parts of the model with S representing the number of periods per season (Debnath and Mourshed (2018)).

[70] Barker, J. (2020). Machine learning in m4: What makes a good unstructured model? *International Journal of Forecasting*, 36(1):150–155

Stationary time series consist of random processes that have constant mean which do not exhibit trend pattern. Augmented Dickey-Fuller Test is performed for this purpose.

In ARIMA models, a model combination benchmark cannot be applied, as there is an infinite number of ARIMA models (Petropoulos et al. (2018)).[71] Also, the running time of ARIMA models is too high in comparison with other models used as a benchmark.

Erdoğdu (2010)[72] used these methods for making predictions of gas demand in Turkey with a data set from the IEA covering quarterly consumption for the period of 1988-2005. For the validation of modelling, the author calculated the absolute value of deviation in million cubic metres from 2000 to 2004 and the annual deviation as a percentage of actual consumption; with the minimum deviation of 1%, the maximum 8.1% and the average at 4%.

[71] Petropoulos, F., Hyndman, R. J., and Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554

[72] Erdoğdu, E. (2010). Natural gas demand in Turkey. *Applied Energy*, 87(1):211–219. https://doi.org/10.1016/j.apenergy.2009.07.006

To evaluate this forecast for 2020, forecast values in billion cubic metres of gas from 2009-2017 are compared with reported data from Eurostat. The average annual deviation expressed as a percentage of actual consumption was at 13.4% , with the minimum deviation in 2011 at 4.9% and the maximum deviation at 23.5% in 2016.[73] Forecasts overestimated gas demand in Turkey for all years. Decisions based on these forecasts would cause overcapacity but no

[73] To complete the analysis, the mean absolute percentage error (MAPE) and the symmetric mean absolute percentage error (sMAPE) are 19.58% and 8.91% respectively.

threat to gas supply in Turkey in any sector (residential, industry and electricity production).

In the R forecast package, the auto.arima() function implements the Hyndman-Khandakar algorithm (Hyndman and Khandakar (2008))[74], which chooses an appropriate model with the lowest Akaike Information Criterion (AIC)

[74] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(1):1–22

$$AIC = -2ln(\hat{L}) + 2p \tag{3.2.13}$$

where p denotes the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model.

4. Statistical/Generalized AutoRegressive Conditional Heteroscedasticity (GARCH) models.

Chkili et al. (2014)[75] used GARCH models for forecasting the conditional volatility and market risk of four commodities: oil, gas, gold and silver. Busse et al. (2012)[76] recommend GARCH and other linear methods for short-term gas price forecasting as their results suggest gas price development is not dominated by nonlinear input factors.

[75] Chkili, W., Hammoudeh, S., and Nguyen, D. K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics*, 41:1–18

[76] Busse, S., Helmholz, P., and Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - an analysis of potential influence factors on basis of a NARX neural network. *Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012*. https://publikationsserver.tu-braunschweig.de/servlets/MCRFileNodeServlet/dbbs_derivate_00027726/Beitrag299.pdf Last accessed 2021-04-14

5. Statistical/Decomposition and trend analysis

There are a few types of decomposition. This work uses the additive and multiplicative model of decomposition based on moving averages (section 4.1) as well as the STL (the Seasonal and Trend decomposition), which handles any type of seasonality. In hybrid models, decomposition serves as a preprocessing step and the modelling of components of decomposition and their residuals can be achieved through an unstructured modelling technique.

In the gas field, Akpinar and Yumusak (2016)[77] applied decomposition, Holt-Winters exponential smoothing and ARIMA on four years of gas consumption data in Turkey (2011-2014), gathered in monthly periods. They concluded that all methods provide satisfying results and differences in accuracy between them are low.

[77] Akpinar, M. and Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies*, 9(9):727

Decomposition introduces simplicity to modelling. Some authors claim the opposite as united data are decomposed into more parts, which goes hand in hand with the natural use of the term complexity as discussed later.

6. Statistical/Linear regression models

Linear regression, as all causal models, is based on the assumption that an output *responds* to the various, up-front chosen variables (predictors). Their number shall be lower than the number of data points and they shall not be correlated.

Precisely, given a set of n measurements with d attributes, each measurement can be represented as a d-dimensional input vector X. Under the assumption, each vector X has an associated scalar property, $s \in \mathcal{R}$; regression analysis derives a functional relationship between each vector X and its associated scalar property s (Rieck (2017)).[78]

[78] Rieck, B. A. (2017). *Persistent Homology in Multivariate Data Visualization*. PhD thesis, Heidelberg University Library. http://archiv.ub.uni-heidelberg.de/volltextserver/22914/1/Dissertation.pdf Last accessed 2021-04-13

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \hat{\beta}_j \tag{3.2.14}$$

The parameter $\hat{\beta}$ is estimated from a training data set $(x_1, y_1)...(x_N, y_N)$, by minimizing the sum of squared residuals (a quadratic cost function corresponding to the log-likelihood of the distribution of the error term) (Hastie et al. (2009)):

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 \tag{3.2.15}$$

Estimation error refers to $\hat{\beta}$ coefficients which vary regarding the true regression coefficients. It is assumed the model is built upon the sample of the data and as the sample N increases, the estimation error decreases (Tashman (2018)). Thus, it gives more weight to outliers in the data set as these have exponentially large residuals. This can be solved by applying the absolute loss function $L(y, f(x)) = |y - f(x)|$ or Huber loss function (below), which squares 'small' residuals and takes the absolute value of them when they are larger (Kuhn and Johnson (2016)).[79]

Notation on prediction values: $f(x) = \hat{y}$

[79] Kuhn, M. and Johnson, K. (2016). *Applied predictive modeling*. Springer, New York, Corrected 5th printing edition. ISBN 978-1-4614-6849-3

$$\mathcal{L}(y, \hat{y}) = (y - \hat{y})^2 \quad ...|y - \hat{y}| \leq \alpha \tag{3.2.16}$$

$$|y - \hat{y}| \quad ...|y - \hat{y}| > \alpha \tag{3.2.17}$$

where y is the target variable, $\hat{y}$ the prediction and $\alpha \in \mathcal{R}^+$ is a hyperparameter.

Multiple linear regression is based on the fact that when all inputs $x_1, x_2, ..., x_p$ are orthogonal (uncorrelated), then their multiple least square estimates $\hat{\beta}_j$ are equal to the univariate estimates. That means, the orthogonality of inputs enables parameter estimates to be independent of each other. For inputs x_1 and x_2, the vector x_2 is regressed on the vector x_1, which leaves the residual vector z. Then, y is regressed on z, which gives the coefficient of x_2. If some inputs are correlated, the residual vector, representing how much of x_p is unexplained by an x_k, would be close to zero, which would make the coefficient $\hat{\beta}_p$ unstable (Hastie et al. (2009)).

Multiple linear regression is used to remove insignificant variables and / or employed to compare and demonstrate performance of other methods, as in Beyca et al. (2019).[80] They compared this method with SVR[81] and ANN methods for monthly forecasts of gas consumption of up to one year in Istanbul.

Similarly to the case study on gas consumption in Germany, Sen et al. (2019) used multiple linear regression to predict Turkey's yearly gas consumption based on socio-economic variables.

[80] Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., and Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, 80:937–949

[81] Support vector regression

THE COMPLEXITY OF THE LINEAR REGRESSION, is measured by *degrees of freedom* – number of parameters chosen by a model (Kuhn and Johnson (2016)).[82]

[82] In econometrics, this approach of measuring complexity has been criticized by Keuzenkamp and McAleer (1997).

7. LOGISTIC MODELLING ANALYSIS and the logistic-population modelling approach assumes there is a maximum demand related to historic extraction / production of all forms of energy resources, which in turn relates to the availability and depletion of the raw materials (Shaikh and Ji (2016)).[83] The following parameters are used: D_{max} - the maximum gas demand a country is expected to achieve in the long term, α - growth parameter, t_{max} - time in years when half of the D_{max} or the $D_{max}/capita$ occur.

   Melikoglu (2013)[84] used the logistic equation for forecasting gas demand in Turkey between 2013 and 2030.

[83] Shaikh, F. and Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, 77:25–32

[84] Melikoglu, M. (2013). Vision 2023: Forecasting Turkey's natural gas demand between 2013 and 2030. *Renewable and Sustainable Energy Reviews*, 22:393–400

8. RANDOM FOREST is a collection of tree predictors $f(x, T, \Theta_k), k = 1, 2, ...K)$ where the Θ_k are independent and identically distributed random vectors. This method has its roots in regression trees and it is constructed by portioning a data set sequentially along values of the explanatory variables (Kaposty et al. (2020)). The stop criterion, i.e. the penalty method, causes the algorithm stop to prevent overfitting. The performance of the model would slightly deteriorate by including extra predictors, as there is a higher chance that a model randomly uses unimportant predictors for splitting (Kuhn and Johnson (2016)).

Breiman (2001)[85] dealt with the problem of overfitting by proposing the mean prediction of many different trees. The method is also used for introducing nonlinearity in the data.

[85] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32. https://doi.org/10.1023/A:1010933404324

The problem of overfitting is typical for *frequentist*[86] methods in contrast to Bayesian approach. In the latter, "the maximization over a subset cannot exceed that over the full set" (Lindley (2001)). Least squares as the method used for fitting "...is equivalent to a Bayesian argument using an improper prio, namely a uniform distribution over the space of the regression parameters" (Lindley (2001)).

[86] Frequentist methods use the error rate form assessing the quality of the model; models based on Bayesian methods use coherence.

9. Temperature correlation models, also called heating degree methods, are used for *residential* demand forecasts only as the idea lies in the strong negative correlation between the daily outside temperature and gas consumption.

   Franco and Fantozzi (2015)[87] suggested forecasting models for residential gas consumption in the winter as follows:

   [87] Franco, A. and Fantozzi, F. (2015). Analysis and clustering of natural gas consumption data for thermal energy use forecasting. *Journal of Physics: Conference Series*, 655:012020

   - an additive model for a total consumption C_t

     $$C(t) = C_N + C_W + C_s + C_r \tag{3.2.18}$$

     where C_N represents the standardized load shapes of production, C_W the weather sensitive component, C_S a random term.

   - a multiplicative model

     $$C(t) = C_N \times f(w) \times f(s) \times f(r) \tag{3.2.19}$$

     where C_N is the base consumption and fs are correction factors for current weather, special events and fluctuation, respectively.

   - a model combining an additive and a multiplicative model

     $$C(t) = F(d(t) \times f(w(t))) + R(t) \tag{3.2.20}$$

   where $C(t)$ is the current consumption at time t, $d(t)$ is the day of the week, $F(d)$ is the daily component, $w(t)$ is a function of the weather data (temperature, humidity and wind chill), $f(w)$ is a weather function and $R(t)$ a term of correction. The accuracy of gas forecasting in temperature correlation models depends on the quality of weather forecasts.

10. Artificial Neural network models are non-linear, non-parametric models enabling the forecasting of any subject without knowledge of the specific relationships between variables. Neural networks can search for their parameters locally and globally (see the distinction in Januschowski et al. (2020)) and almost all of them work without any non-convex loss function.

    Šebalj, Dario and Dujak Davor and Mesaric Josip (2017) concluded neural networks are currently the most commonly used models in gas forecasting.

THE ISSUE OF "ADAPTING WEIGHTS" in a neural network can be seen as a special case for estimating parameters of any other functional model (Werbos (1988)).[88] Concretely, a coefficient β_{jk} is the effect of the j_{th} predictor on the k_{th} hidden unit (Hastie et al. (2009)). In the first step, data normalization is required as normalization causes the higher population of data for the same manifold space.

[88] Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356

ACTIVATION FUNCTION links the data in predictors with the first hidden layer.[89] There are several commonly used activation functions, the ReLU (rectified linear unit) $\sigma(a) = max\{0, a\}$ is currently the most popular, while the sigmoid function $\sigma(a) = \frac{1}{1+e^a}$ has been used in the more traditional approach. Both functions are advantageous over the threshold activation (e.g. $\sigma(z) = 1$ if $z > 0$ and 0 if otherwise) as they can be trained using gradient based methods (Livni et al. (2014)).[90] *Stochastic* gradient descent involves random shuffling of the training data set before each iteration; this enables different orders of updates to the model parameters. The linear activation function is seen as problematic; the output of the second layer is just a different linear function of the first layer as was shown in the appendix of Barker (2020).

[89] Single-hidden layer neural networks are an example of the learning method with additive expansion, more details on additive models in Hastie et al. (2009), page 341.

[90] Livni, R., Shalev-Shwartz, S., and Shamir, O. (2014). On the computational efficiency of training neural networks. `https://arxiv.org/pdf/1410.1141` Last accessed 2020-12-20
Polynomial networks are networks using the squared activation function $\sigma_2(x) = x^2$.

IN SUPERVISED (STRUCTURED) LEARNING (previous models), input $X^T = (X_1, ..., X_p)$ and output vectors $Y = (Y_1, ..., Y_m)$ are specified and the network tries to minimize an error, characterized by a loss function $L(y, \hat{y})$, for a known set of target values (answer $\hat{y}_i$ for each x_i); in unsupervised learning, targeted output data is not required and this increases degrees of freedom, which can result in overfitting. As Barker (2020) puts it; "the search space of the model increases exponentially with the degrees of freedom of the model, but the number of points to fit only increases linearly with the size of the data set."

In unsupervised learning, there is a set of N observations $(x_1, x_2, ...x_N)$ of a random p-vector X having joint density $Pr(X)$. Properties of this probability density are deduced without providing any targeted output data for each observation (Hastie et al. (2009). A learning is understood as a density estimation problem if we suppose that (X, Y) are random variables represented by joint probability $Pr(X, Y)$).

The number of parameters p is counted according to the formula:

$$p = H(P + 1) + 3 + 1 \tag{3.2.21}$$

where H denotes the number of hidden layers and P the number of predictors (inputs).

FORECASTING IS AN EXTRAPOLATION PROBLEM, and unstructured models are optimal for interpolation, as the relationship between forecasts and its lags is more flexible (Barker (2020)). In unstructured models, the curse of dimensionality is solved by assuming the manifold, hypothesis, i.e. data in a data set are discrete samples of a continuous manifold of some dimension. The manifold hypothesis can also be seen as one of the tools for reducing the complexity. There is no single algorithm for verifying the hypothesis, but manifolds enable a smooth mathematical structure, especially while analyzing natural phenomena (Rieck (2017)).[91] In the best possible case, the manifold where the data lie should be as low-dimensional as possible and as densely populated as possible (Barker (2020)).

[91] Rieck, B. A. (2017). *Persistent Homology in Multivariate Data Visualization*. PhD thesis, Heidelberg University Library. `http://archiv.ub.uni-heidelberg.de/volltextserver/22914/1/Dissertation.pdf` Last accessed 2021-04-13

Any low-dimensional entity can be understood as a subset of the higher dimension; this is demonstrated in the two pixeled pictures of identical size. The right side of the merged figure 3.5[92] depicts the result of using a random pixel generator with a pixel size of 4, the left side shows one possible subset out of many. The left configuration of colours of pixels of the same size is typical for what is identified as "Marlene Dietrich" and any picture of Marlene Dietrich's portrait shows similarities in the configuration of colours of pixels with the picture on the left. These configurations evoke associations with the Boltzmann's formula for entropy S

[92] Sources: Random Pixel Generator: `http://pixelmonkeys.org/`, with a pixel size of 4. The left side of the image, "marlene dietrich top hat, for morocco" by carbonated, is licensed with CC BY-NC-SA 2.0. To view a copy of this license, visit `https://creativecommons.org/licenses/by-nc-sa/2.0/`. The online tool used for the pixelated version: `https://onlinepngtools.com/pixelate-png`.

$$S = k_B log W \tag{3.2.22}$$

where $k_B = 1.3807x10^-23J/K$ denotes Boltzmann's constant, the conversion factor between units of temperature and units of energy and W the number of real microstates corresponding to the gas's macrostate. The right part of the figure depicts microstates of presumably equally probable assigning of any colour from the range white-black to a pixel. The left figure does not fulfill this condition; there is a smaller choice of assigning colors to a pixel for producing a picture resembling Marlene Dietrich, and thus it is a subset of the picture on the right.

Manifold hypothesis heuristically explains why machine learning techniques work. A model needs to focus on a few key features in a data set to make decisions. The task shall be as specific as possible with lots of data enabling to find these features.

IN THE BACKPROPAGATION ALGORITHM, the cost function tries to minimize the error, i.e. the sum of the squared residuals by back propagating to the hidden layer, and weights are either increased or decreased until the desired output is achieved. This was the first algorithm that allowed the adaptation of *all weights* of a neural network. As shown in figure 3.6, a negative partial derivative will increase the weight and a positive partial derivative will decrease it until the local minimum is found (Günther and Fritsch (2010)).[93]

[93] Günther, F. and Fritsch, S. (2010). Neuralnet: Training of neural networks. *R Journal*, 2(1):30–39. `https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf` Last accessed 2020-12-20

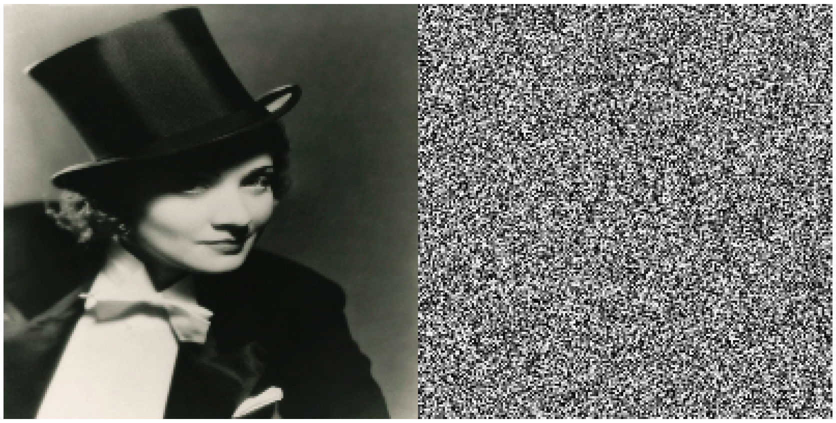


Figure 3.5: Visualisation of relation of low dimension to high dimension.

In contrast to classical back-propagation, *resilient* back propagation enables the change of the learning rate η_k during the process of searching for the minimum. In shallow areas, for speeding up the convergence, the learning rate η_k will increase "if the corresponding partial derivative keeps its sign" (Günther and Fritsch (2010)), (see the second formula with the sign below). Changing the sign will cause the learning rate η_k to slow down as it means that the minimum has been missed. For comparison,

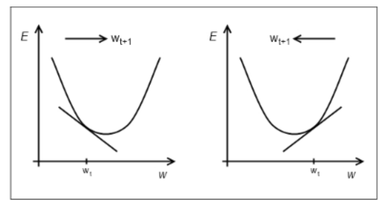


Figure 3.6: Basic idea of the back-propagation algorithm illustrated for a univariate error function E(w) as in Günther and Fritsch (2010).

- the rule for adjusting weights in classical backpropagation:

$$w_k^{(t+1)} = w_k{}^t - \eta \frac{\partial E^{(t)}}{\partial w_k{}^{(t)}} \tag{3.2.23}$$

  where t indexes the iteration steps and k the weights.

- the rule for adjusting weights in resilient backpropagation:

$$w_k^{(t+1)} = w_k{}^t - \eta sign\left(\frac{\partial E^{(t)}}{\partial w_k{}^{(t)}}\right) \tag{3.2.24}$$

Learning rate η_k should decrease to zero, as the iteration r approaches infinity.

The neuralnet package, tested in the modelling section, uses classical propagation as described here, resilient backpropagation with or without weight backtracking and the modified globally convergent version.

Weight backtracking adds a smaller value to the weight in the next step. The technique prevents the algorithm from jumping over the minimum by undoing the last iteration (Günther and Fritsch (2010)).

Raza and Khosravi (2015)[94] identified four learning issues related to this algorithm: getting trapped in local minima instead of the global minimum i.e. there could be another set of parameters uniformly better, network paralysis, temporal instability and lack of generalization of the network, resulting in overfitting, characterized as unintended memorization of synaptic weight values. To avoid overfitting, several approaches have been developed, such as early

[94] Raza, M. Q. and Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372

stopping (the procedure will stop before approaching the global minimum; when an error estimate starts to increase) or weight decay – adding a penalty to the error function, with λ as its tuning parameter. For λ estimation, the cross-validation is used (Hastie et al. (2009)).

FOR SHORT-TERM LOAD FORECASTING, hybrid models have been proposed such as ANN with fuzzy and genetic algorithm, ANN with wavelet and time series, ANN with genetic algorithms etc. Examples are to be found in Raza and Khosravi (2015).

11. SUPPORT VECTOR MACHINES (SVMs)

Originally, Cortes and Vapnik (1995)[95] developed SVMs for data classification; the main idea is depicted in figure 3.7. To make boundaries more flexible, the feature space is enlarged by using polynomials (expansions).

The optimal hyperplane is denoted as:

$$w_0 \cdot z + b_0 = 0 \tag{3.2.25}$$

The weights w_0 could be written as the linear combination of support vectors:

$$w_0 = \sum_{support \quad vectors} \alpha_i z_i \tag{3.2.26}$$

where α_i are weights of the output units and z_i weights from a hidden unit. As it can be seen from a formula, new samples enter the model as the sum of inner products so the formula can be rewritten with the kernel function K(·). Various kernel functions can be chosen: e.g. linear, polynomial, or the radial basis function. The tuning parameter sigma impacts the smoothness of the decision boundary (Kuhn and Johnson (2016)); its underestimation would cause the boundary to be too sensitive to noise.

The cost parameter, (i.e. setting the price for misclassified samples in the training set), is attached to residuals, not to the parameters. It can be manually set or determined by using cross-validation. With large cost parameters the model becomes flexible and likely overfits. Thus, this parameter can be understood as a measure of complexity for SVM.

SVMs are one of the methods for performing regression analysis in machine learning; the support vector regression (SVR) is a (linear) regression formulation of the support vector machines. Considering the linear regression model (Hastie et al. (2009)):

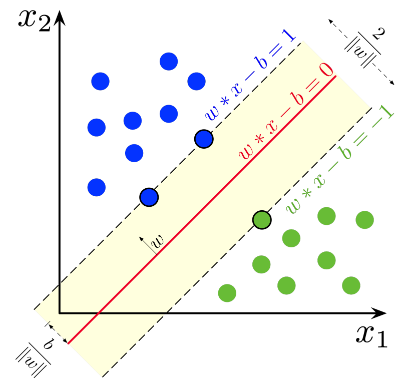


Figure 3.7: An example of how to separate training data in a two-dimensional space. The support vectors define the margin of largest separation between the two classes. Source: Larhmam (`https://commons.wikimedia.org/wiki/File:SVM_margin.png`), „SVM margin", `https://creativecommons.org/licenses/by-sa/4.0/legalcode`

[95] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297

$$f(x) = x^T\beta + \beta_0, \tag{3.2.27}$$

β is estimated by considering minimization of

$$H(\beta, \beta_0) = \sum_{i=1}^{N} V(y_i - f(x_i)) + \frac{\lambda}{2}\|\beta\|^2 \tag{3.2.28}$$

where

$V_{\epsilon(r)} = 0 \quad$ if $|r| < \epsilon$, and

$V_{\epsilon(r)} = |r| - \epsilon, \quad$ otherwise.

From the formula above, it is clear that data points with residuals within threshold ϵ are ignored in the regression fit, analogically to the SVM in the classification problem, where points on the right side of the division and far away form it are also ignored.

12. HYBRID MODELS link causal methods (i.e. input variables are assumed to affect the output) and data-based models. They are composed of structured and unstructured methods, e.g. using machine learning for forecasting the trend component of the decomposition only and are regarded as the next and further developed form than ensemble models. However, the choice of models to be hybridized remains arbitrary. Since they have become a popular method, the forecasting community raised the following questions:

   - Under which conditions does a model combination perform better?
   - Would decision makers accept accuracy risk[96] (a higher risk of poor forecasts) while combining models?
   - Do more complex models gain outperform simple models when the criterion of the complexity is measured by computing time?[97]

   de Nadai and van Someren (2015) tested the combination of the ARIMA and ANN model to detect anomalies in gas consumption.

13. ENSEMBLE OF MODELS[98]

   One single model with a single benchmark does not address model, data, or parameter uncertainty.[99] While combining models, there is a chance a new model will provide information that has not been caught by other models. Thus, a combination of models keeps the bias about the same but decreases the variance (Atiya (2020)). While constructing an ensemble, a practitioner answers two questions:

[96] The indication of the accuracy risk is provided by the variance; "how likely a very poor forecast is for a given series" (Lichtendahl and Winkler (2020)).

[97] In this section, we do not explore questions 1 and 2 further. Question 3 is discussed in the section on complexity.

[98] This method was preferably used in the *M4* Competition, organized by the International Institute of Forecasters.

[99] Uncertainty types are not of the same magnitude; the effect of parameter uncertainty seems to be of a smaller order of magnitude than that due to other sources (Chatfield (2001)).

- Which and how many forecasting models are to be combined. A method with poor forecasts contributes to the diversity of the pool of models and therefore it shall not be left out. On the other hand, one of two poor performing methods with highly positively correlated errors can be excluded due to no gain in diversity (Lichtendahl and Winkler (2020)).[100]
- Which method to use for their combination and how to choose the best weights. Models are aggregated by using mean (the preferred method), median, mode, or trimmed mean. Gaba et al. (2017)[101] suggests combining between five and ten forecasts for optimal results. The greater diversity in the pool, i.e. with minor direct relation,[102] the better results.

[100] Lichtendahl, K. C. and Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, 36(1):142–149. https://doi.org/10.1016/j.ijforecast.2019.03.027

[101] Gaba, A., Tsetlin, I., and Winkler, R. L. (2017). Combining interval forecasts. *Decision Analysis*, 14(1):1–20

[102] Among forecasting methods, diversity is measured in terms of correlations among their forecasting errors (Lichtendahl and Winkler (2020)).

In an analogy to complex systems, a practitioner acquires more individual predictions (prediction trajectories) by suitable perturbation of the initial state of the system, to which complex systems are sensitive (Pelikán (2014)). In a similar fashion, various models and their functions fit the data and produce forecasts ("prediction trajectories").

3.2.6 *Excluded models*

Equilibrium- and linear optimization models

Computable general equilibrium (CGE) models

Equilibrium models (e.g. MARKAL, IKARUS, TIMES) assess technology options based on the simple decision rule of cost optimisation (Bale et al. (2015)).[103] In these models, agents (e.g. households, energy trading companies) are assumed to act as rational economic actors with the ability of perfect foresight. Scenarios are unique for each case; only their full description captures the essence. The next example illustrates the point. Eser et al. (2019) and Gillessen et al. (2019) share similar objects and time frames in their studies: the impact of the Nord Stream 2 pipeline and liquefied natural gas (LNG) on gas trade, the security of supply up to 2030 and infrastructure expansion. However, whereas Eser et al. (2019) assume gas from production countries is annually and hourly limited, Gillessen et al. (2019) examines energy security by modelling the gas flow disruption in one of the transmission pipelines and models the gas flows and velocities of the remaining parts of the transmission system.

[103] Bale, C. S., Varga, L., and Foxon, T. J. (2015). Energy and Complexity: New ways forward. *Applied Energy*, 138:150–159. https://doi.org/10.1016/j.apenergy.2014.10.057

The study by Eser et al. (2019) simulates a high-pressure gas network with 500 network nodes (i.e. points where the gas flow is measured), and 150 compression stations and gas storage sites within the borders of Germany to identify bottlenecks in the gas system. Input on hourly imports into Germany is produced by using a Monte Carlo approach: "30 optimizations of the gas sourcing, each with stochastically varied gas prices at the boundaries of the network, are solved."[104] Thus, gas imports represent an optimized input, meaning the least-cost gas imports' mix from probable production countries and world LNG price, and assuming knowledge of events during the simulated year. Authors validate the novel simulation at the annual level although hourly values were produced first.

[104] Eser, P., Chokani, N., and Abhari, R. (2019). Impact of Nord Stream 2 and LNG on gas trade and security of supply in the European gas network of 2030. *Applied Energy*, 238:816–830. https://doi.org/10.1016/j.apenergy.2019.01.068

THE IEA AND THE EIA[105] work with macroeconomic models with an outcome in the form of scenarios. The IEA applies the World Energy Model (WEM), a partial equilibrium simulation model covering global energy supply, transformed energy and its demand. The energy demand / supply equilibrium is computed for the minimal total cost of providing energy services. World Energy Outlook scenarios are taken as inputs for other models, such as projections on the energy security (Ang et al. (2015)).[106] The Annual Energy Outlook of the EIA is produced by an energy-economic model of the US energy system, the "National Energy Modeling System," based on the two main drivers of GDP and energy intensity (Bezdek and Wendling (2002)).[107]

[105] The US Energy Information Administration.

[106] Ang, B. W., Choong, W. L., and Ng, T. S. (2015). Energy security: Definitions, dimensions and indexes. *Renewable and Sustainable Energy Reviews*, 42:1077–1093

[107] Bezdek, R. and Wendling, R. (2002). A half century of long-range energy forecasts: Errors made, lessons learned, and implications for forecasting. *Journal of Fusion Energy*, 21. https://doi.org/10.1023/A:1026208113925

RETROSPECTIVELY, LARGE DEVIATIONS FROM THE FORECASTS and following wrong policy assumptions have been observed. At first sight, this is not obvious as low errors for total energy consumption conceal much larger errors in sectors that offset each other when aggregated (Winebrake and Sakva (2006)).[108] They confirmed intuitive assumptions: a) forecasts exhibit increased uncertainty when time horizons are lengthened[109]; and b) certain sectors (e.g. residential) demonstrate a more accurate level of forecasting than others (e.g. transport).

[108] Winebrake, J. J. and Sakva, D. (2006). An evaluation of errors in US energy forecasts: 1982–2003. *Energy Policy*, 34(18):3475–3483

[109] Magdowski and Kaltschmitt (2017) analyzed day-ahead power forecasting from wind turbines in Germany and came to the same conclusion. Moreover, the prediction accuracy depends on the predicted feed-in volume, with small feed-in volumes decreasing the accuracy.

In forecasting, statistic properties of data are checked to choose proper models for producing forecasts, and so assumptions on the future development of an energy system are not needed.

ENERGY-SECTOR SPECIFIC ERRORS include

- underestimation of oil and gas production,
- projecting exhaustion of energy resources. Prediction of oil peak in the time horizon of 10-15 years from the year of estimation,
- overestimation of energy consumption,
- an assumption that "technically feasible technologies or technologies feasible in an engineering sense will penetrate the market in the future"(Bezdek and Wendling (2002)).

THE PRIVATE OIL AND GAS COMPANIES Shell, ExxonMobil, Statoil, and the BP (BP (2019)) publish their own projections. Outcomes of all equilibrium models do not equal forecasts as no value is assigned to the probability of various scenarios. Also, they do not grasp the complex reality of real processes in the economy/energy sector (Bale et al. (2015)). However, they reflect the storyline of an expected future scenario as it was shown with the TIMES model for the case study Portugal in Fortes et al. (2015).[110]

[110] Fortes, P., Alvarenga, A., Seixas, J., and Rodrigues, S. (2015). Long-term energy scenarios: Bridging the gap between socio-economic storylines and energy modeling. *Technological Forecasting and Social Change*, 91:161–178

3.2.7 Econometric models, grey models, genetic algorithms

Econometric forecasting models use the Cobb-Douglas formula for gas demand in log-linear form:

$$Ln(Gas \quad demand) = \gamma + \alpha Ln(Price) + \beta Ln(GDP \quad per \quad capita) \tag{3.2.29}$$

where α denotes elasticity of price, β elasticity of GDP per capita and γ is a constant (Dey et al. (2011)).[111] These models have been omitted as their research questions relate more to price and income (in)elasticities of demand for gas than to the actual forecasting.

[111] Dey, H. S., Kabir, M. A., Wadud, Z., Khan, S. I., and Azad, M. A. K. (2011). Econometric modeling and forecasting of natural gas demand for power sector in Bangladesh. IEEE Region 10 Annual International Conference, Proceedings/TENCONISBN 978-1-4577-0256-3

Second, grey prediction models were taken out of the list. Ma and Li (2010) used a bivariate model of grey system $GM(1,2)$ between gas consumption and GDP in China for the yearly prediction of gas consumption from 2009 to 2020. The following table 3.3 compares their prediction with data from the Statistical Review of World Energy (BP); the same source of information was used in their paper. From 2015 on, the model over-predicts gas consumption in China.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Prediction	89.9	104.2	120.8	140.0	162.2	188.0	217.9	252.6	292.7	339.3
BP, Statistics	90.2	108.9	135.2	150.9	171.9	188.4	194.7	209.4	240.4	283.0

Table 3.3: Comparison of predictions made by Ma and Li (2010) with the statistical data from the BP, Statistical Review of World Energy 2019, Gas consumption in billion cubic metres.

Third, genetic algorithms (GA) are stochastic optimization algorithms that supposedly simulate biological evolution (reproduction, mutation, recombination, and selection). It is a class of population-based algorithm and finds the optimal solution on the basis of the optimal point of a population (Raza and Khosravi (2015)).[112] Genetic algorithms can often be used for finding optimal tuning parameters in the hybrid methods.[113]

Although the last two models contribute to the diversity of models tested on gas forecasting, they are marginal. Therefore, they were taken out of the list of models to consider.

[112] Raza, M. Q. and Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372

[113] Yu and Xu (2014) combined GA with improved BP neural network to form a new prediction model.

3.2.8 Conclusion

To conclude this chapter, a practitioner can use a) a single already established model, b) create an ensemble of well-established models, or c) if there is a data scientist in a team, the team's own model or an ensemble of known models. One method is set as a benchmark, for example the persistence model or the Comb Method (the arithmetic average of the Simple, Holt and Damped exponential smoothing models.[114] Simple exponential smoothing methods are used when data has no trend or seasonal pattern. In these methods, weights exponentially decrease for past data; algorithms select the best in-sample models based on information criteria such as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion) before forecasts are generated.

[114] Holt linear (additive trend component with no seasonal component (A,N), Damped model (Additive damped trend component with no seasonal component, (A_d, N)) and Holt-Winters methods (both components; trend and seasonal additive, (A,A)) are three out of 15 types of exponential smoothing methods according to Hyndman and Khandakar (2008). Combinations are created for seasonal component (N-none, A-additive, M-multiplicative) and for trend component (N - none, A - additive, A_d - additive Damped, M - multiplicative, M_d - multiplicative Damped.

The forecasting methods above do not require domain knowledge (the knowledge of a time series' commodity or sector). As Darin and Stellwagen (2020)[115] point out: "if a business series exhibits unusual structural changes, the reason is virtually always known (e.g. a patent expired, the company merged with another firm, etc.) and this domain knowledge will guide how the series should be forecast." In gas forecasting, authors applying domain knowledge do not publish their results in forecasting magazines.[116] Although gas forecasting is part of business forecasting, one's knowledge of the energy market at city, country or global level cannot reach the depth of a company's insider knowledge.

[115] Darin, S. G. and Stellwagen, E. (2020). Forecasting the M4 competition weekly data: Forecast pro's winning approach. *International Journal of Forecasting*, 36(1):135–141

[116] Franco and Fantozzi (2015), one of the few to do so, analysed and clustered gas consumption data for thermal energy use forecasting in Italy.

Forecasts shall be computed fast; there may be a trade-off between the time spent computing and the accuracy of forecasts. The final check on whether forecasts make sense can take a few hours or a few working days. Instead of getting point forecasts right, prediction intervals, i.e. getting the tails (upper quantiles) right matter in business and energy forecasting. Then, error measurement in predicting prediction intervals is understood as the difference between expected and observed coverage. Here one needs to differentiate between conditional (ex post[117]) where it is assumed true values of variables are known and unconditional (ex ante) forecasts that reflect the uncertainty of future variables. As for regression analysis, since the random error and the estimation error does not change with time it is the uncertainty of inputs that causes increases with time. Thus, in the ex ante forecasting, the width of prediction intervals shall increase; Tashman (2018) suggests to double the width.

In the exponential smoothing, prediction intervals might be used for identifying outliers to remove them and then the model selects the best model in terms of AIC or BIC criterion to produce forecasts.

[117] Forecasts computed in the modelling section are of the ex post nature if not stated otherwise.

The amount, quality of time series used and contextual information about them also determine the performance of chosen models. For example, if dates for the series are not present (as e.g. in the M4 forecasting competition), cross-series learning, which is crucial for getting benefits from unstructured models, cannot be applied.

Study cases in the modelling section represent a real-world application of forecasting principles as a) dates in a data set are available b) the series are related to one another.

The scale and its economy influence the predictability of data; the gas demand of the local distribution company is easier to predict than the demand of one single house.

3.3 Complexity

3.3.1 Terminology

"Some folk love complexity for it hides inadequacies and even errors." (Lindley (2001))

Explaining the role of complexity in forecasting is by itself a complex issue. Authors of textbooks count on reader's contextual understanding of complexity; a few textbook examples can be found in Applied Predictive Modelling by Kuhn and Johnson (2016):[118]

fewer predictors decrease complexity,
tuning parameters control the complexity of the model,
degrees of freedom measure complexity for linear regression models,
SVM and random forests are associated with higher complexity,
and finally
"all other things being equal, simplicity is favored over complexity."
Is the same definition of complexity implied in each case?

[118] Kuhn, M. and Johnson, K. (2016). *Applied predictive modeling*. Springer, New York, Corrected 5th printing edition. ISBN 978-1-4614-6849-3

Support Vector Machines

First, the term is not universally defined. The natural use of both terms, complexity and simplicity, does not point to a direct antonymous relationship. The Oxford Learners Dictionary of Academic English defines complexity as 1) the state of being formed of many parts; the state of being difficult to understand 2) for a plural form – the features of a problem or situation that are difficult to understand. Simplicity is defined as 1) the quality of being easy to understand and use, 2) the quality of being natural and plain, 3) in plural: an aspect of something that is easy, natural or plain. These definitions do not draw a boundary line; the ease of understanding depends on one's intellect and experience. Still, complexity appears to be related to any subject formed of many parts.

Herbert A. Simon in (Zellner et al. (2002))[119] defined complexity by the length of the string of zeros and ones; and simplicity as the reciprocal of its complexity. In similar fashion, he measured complexity and simplicity of the formula used to describe the string, more generally: the data set. Then, parsimony is a relation between strings representing a data set and the other representing a formula for that set. More concretely, it is "the ratio of the complexity of the data set to the complexity of the formula (Zellner et al. (2002))." Any formula involving a pattern found in the data set would be of decreased length and thus of a decreased level of complexity.

"Consider a data set represented as a one-dimensional ordered string of 0s and 1s. By the complexity of this string, I will mean simply its length, and by its simplicity, the reciprocal of its complexity." Herbert A. Simon

[119] Zellner, A., Keuzenkamp, H. A., and McAleer, M. (2002). *Simplicity, inference and modeling: Keeping it sophisticatedly simple*. Cambridge University Press, Cambridge and New York. ISBN 0521803616

The first approximation to measuring complexity can be calculated by a measure suggested by Jeffreys in Zellner et al. (2002):

$$C = Order + Degree + S \qquad (3.3.1)$$

where S stands for the Sum of absolute values of normalized coefficients. The last variable does not equal often-cited numbers of parameters in the model, it is the sum of numbers of parameters times their absolute values (Keuzenkamp and McAleer (1997)).[120]

[120] Keuzenkamp, H. A. and McAleer, M. (1997). The complexity of simplicity. *Mathematics and Computers in Simulation*, 43(3-6):553–561

In forecasting, journal articles identify models of high accuracy (based on the test set) without pointing out "why this method outperforms all the others or why a model A outperforms a model B?" Complexity may partially explain it. However, whereas models are relatively easily compared based on the pre-chosen criteria of their predictive performance, quantification measures cannot be developed for their structural differences that might be able to explain the question above. Structural differences remained ignored in the literature (Rieck (2017)).[121]

[121] Rieck, B. A. (2017). *Persistent Homology in Multivariate Data Visualization*. PhD thesis, Heidelberg University Library. `http://archiv.ub.uni-heidelberg.de/volltextserver/22914/1/Dissertation.pdf` Last accessed 2021-04-13

Coming back to quantifiable criteria, in the simplest terms, complexity equals the computing time necessary for a forecasting model to complete the task of forecasting. Thus, in some forecasting competitions, such as the M4 Forecasting competition, this was used as a measure of time efficiency. In general terms, this notion is labelled as computational complexity.

The next section continues with discussing quantifiable aspects of complexity only.

In mathematics, quantifying regularity is related to the concept of entropy, since entropy is associated with randomness as the measure of irregularity and disorderliness. Thus, complex and random systems produce larger entropy and vice versa (Li and Zhang (2008).[122] Tools include Multi-scale entropy, the Kolmogorov-Sinai entropy to classify *deterministic* dynamical systems, the sample entropy developed by Richman and Moorman (2000) and few others. Often they do not share the same definition of entropy, or their assumptions for its application differ.

[122] Li, Z. and Zhang, Y.-K. (2008). Multiscale entropy analysis of Mississippi river flow. *Stochastic Environmental Research and Risk Assessment*, 22(4):507–512. `https://doi.org/10.1007/s00477-007-0161-y`

Kolmogorov-Sinai entropy h, in the ergodic theory[123], is a quantitative measure of the impossibility of perfect forecasts (Zambella and Grassberger (1988)).

"Assume we have observed a system with some measuring device Γ (which has some finite resolution), at all times $\leq t$. Then, we cannot completely predict what the same device Γ will measure at $t + \tau$, $\tau > 0$. Instead, there will be a gap $\leq h\tau + 0(\tau^2)$ between the information obtainable with Γ and the information of the best forecast based on measurements with Γ." (Zambella and Grassberger (1988)). This idea of entropy, as a rate of information generation of chaotic system, was

[123] Ergodic theory studies ergodicity. "In an ergodic system, the trajectory of almost every point in phase space eventually passes arbitrarily close to every other point (position and momentum) on the surface of constant energy." (Sethna, P. James (2020)). The phase space is the 6N-dimensional space (P, Q) where P stands for momentum and Q for position. Sethna, P. James (2020) states our inability to prove that systems in the research are ergodic.

modified by Grassberger and Procaccia in 1983 to be able to calculate such a rate from time series data.

3.3.2 *Approximate Entropy and Sample Entropy*

Relative measure for comparing data sets

APPROXIMATE ENTROPY (APEN) quantifies the regularity in a data set with at least $1,000$ data points; the larger value of ApEn denotes greater randomness, and a smaller value corresponds to more cases of patterns in the data (Pincus (1991)).[124] Given the data set of N data points, "$ApEn(m, r, N)$ is approximately equal to the negative average natural logarithm of the conditional probability that two sequences that are similar for m points remain similar (within a tolerance r) at the next point" (Ramesh (2011)) whereas

- m is the length of sequences compared and
- r is a tolerance window (that is, tolerance for accepting matches).

As Pincus (2008) points out, we can imagine "partitioning the state space into uniform width boxes of width r, from which we estimate m_{th} order conditional probabilities. ApEn is unaffected by noise of magnitude below this filter level r and is finite for both stochastic and deterministic processes, in contrast to K-S entropy.[125] However, ApEn lacks relative consistency and depends on the length of time series.

The tolerance is set to the inner product of rDo and standard deviation of the data set to allow comparison on data sets with different amplitudes. Usually, data are normalized to have standard deviation equal to one (Richman and Moorman (2000)).

SAMPLE ENTROPY (RICHMAN AND MOORMAN (2000))[126] IS THE MODIFICATION of the approximate entropy as it searches for any repeated patterns of various lengths but does not count self-matches. This was justified by understanding entropy as the rate of information production (Chen (2002)[127]) and in this context, comparing data with themselves does not make sense (Richman and Moorman (2000)). If B denotes the total number of matches of length m and A is the total number of forward matches of length $m + 1$, then with the notation:

$$A/B = [A^m(r)]/[B^m(r)] \tag{3.3.2}$$

Sample Entropy can be expressed as

$$SampEn(m, r, N) = -ln(A/B) \tag{3.3.3}$$

APPROXIMATE ENTROPY AND SAMPLE ENTROPY increase as a data set increases.

The Sample and Approximate Entropy was computed with the length of reconstructed vector 1 and 2 (for chaotic processes) and r according to the recommended formula $r = 0.2 \cdot sd$(time series) for gas import

[124] Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6):2297–2301

[125] As Pincus (2008) puts it: ""Descriptively, ApEn thus measures the logarithmic likelihood that runs of patterns that are close (within r) for m contiguous observations remain close (within the same tolerance width r) on next incremental comparisons. Opposing extremes are perfectly regular sequences, (e.g. sinusoidal behavior, very low ApEn), and independent sequential processes (very large ApEn)."

[126] Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology. Heart and circulatory physiology*, 278(6):H2039–49

[127] Chen, J. (2002). An entropy theory of value. *SSRN Electronic Journal*. `http://dx.doi.org/10.2139/ssrn.307442`

values for Germany, from the data set used for forecasting gas imports, with the results displayed in table 3.4. A low value of entropy indicates a time series is either deterministic and easier to predict; a higher value indicates randomness. Sample Entropy does not depend *that much* on the length of the time series.

Table 3.4: ApEn and Sample Entropy for the data set "Gas imports"

Embedding dimension	Approximate Entropy	Sample Entropy
1	1.635595	-
2	1.037104	1.884978

3.3.3 Complexity of forecasting models

General associations of complexity in the literature on predictive modelling

In statistical learning, complexity of methods is understood as the number of degrees of freedom (also referred to as the number of free parameters in the model), "difficulty in estimation," or as the number of variables. It is specified for every forecasting method. In the section *Related work*, typical accuracy criteria for measuring out-of-the sample errors are listed; here we discuss training errors and their complexity. Again, let Y denote the output (a target variable), X vector of inputs and $\hat{f}(X)$ the prediction model from a training data set τ. The loss function is denoted by $L(Y, \hat{f}(X))$ as in Hastie et al. (2009). The training error is the average loss of the training sample (Hastie et al. (2009)):

Ghandar et al. (2016) measured the model complexity in two aspects: a) the number of rules within a rule-base, and 2) the number of inputs utilized within the rule-base. They concluded "the use of overly complex models leads to a greater reduction in performance in a dynamic environment with time-variant relationships than in a static environment where causal relationships do not change substantially."

$$err = \frac{1}{N}(\sum_{i=1}^{N} i = L(y_i, \hat{f}(x_i)) \qquad (3.3.4)$$

With higher number of degrees of freedom (for non-parametric methods), the training error decreases as the function of a model describes the data fully, but this overfitting issue results in poor performance of a model for the test-sample. Methods marked as *stable* are supposed to have the same in-sample and out-of-sample errors.

The Akaike Information Criterion is a penalized method based on the in-sample fit (Hyndman and Khandakar (2008)).[128]

$$AIC = L^*(\hat{\theta}, \hat{x}_0) + 2p \qquad (3.3.5)$$

[128] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(1):1–22

where p is the number of parameters in θ plus the number of free states in x_0, and $\hat{\theta}$ and $\hat{x}_0$ are the estimates of θ and x_0. Since the AIC is based on likelihood, the criterion enables the choice of identical point forecasts from two models (Hyndman and Khandakar (2008)).

Green and Armstrong (2015)[129] defined simplicity as "processes that are understandable to forecast users." Users, not researchers. In this respect, naïve or no-change models without seasonal adjustment, seasonal adjustment, single-exponential smoothing, Holt's exponential smoothing, dampened exponential smoothing and simple average of the exponential smoothing forecasts are regarded as simple. They conclude that "complexity beyond the sophisticatedly simple fails to improve accuracy in all but 16 of the 97 comparisons in 32 papers that provide evidence"(Green and Armstrong (2015)).

[129] Green, K. C. and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8):1678–1685

Although it is not the goal of this work to prove that in every case higher complexity will not improve the accuracy of the model, the research *in other fields* does go in this direction, e.g. Andrade-Cabrera et al. (2018) found they can reduce the model complexity, defined through computational tractability, by half for Urban Building Energy Modelling (UBEM) and still retain annual energy estimation errors below 10 per cent for a single building.

Hyndman and Kostenko Andrey V. (2007) point out the misconception that ARIMA models are more complex than Holt-Winters models and thus need a larger data set. In their article[130] it is shown that ARIMA models actually need the minimum number of 16 observations for estimating a seasonal ARIMA model (monthly data), while for the Holt-Winters model it would be 17 observations. With these numbers of observations, prediction intervals would still be finite.

[130] Hyndman, R. J. and Kostenko Andrey V. (2007). Minimum sample size requirements for seasonal forecasting models. *Foresight*, (6):12–15. https://robjhyndman.com/papers/shortseasonal.pdf

For support vector machines, while developing the algorithm, the order of operations has been interchanged. To start, two vectors are compared in the input space and a non-linear transformation takes place while computing the value of the result. Thus, the complexity of the model does not depend on the dimensionality of the feature space, but on the number of support vectors (Cortes and Vapnik (1995)). Changing the parameter C enables the trade-off between the complexity of decision rule and frequency of error.

As for neural networks, let the sample complexity be defined by the number of examples required to learn the class, i.e. the set of all prediction rules obtained by using the same network architecture[131] while changing the weights of the network. Until now, theoretical work on neural networks has not been promising, still in practice neural networks yield results thanks to a few tricks used. Livni et al. (2014) listed among them the change of the activation function, over-specified networks and the regularization of weights to speed up the convergence. Changing the activation function to the squared function $\sigma_2(x) = x^2$ makes sample complexity grow (the increase of sample complexity is caused by increasing the depth of the network).

[131] The number of layers of the network characterizes the depth of the network. The total number of neurons represents the size of the network (Livni et al. (2014)).

3.3.4 *Complexity of a phenomenon in relation to modelling*

If the process representing the forecast phenomenon is of dynamic (complex) nature, the complexity of this process may influence the choice of a forecasting model. A complex system is a system that consists of many mutually interacting components, and that shows emergent behaviour, i.e. the collective behaviour evidences some traits that cannot be easily derived or explained based on behaviour of individual parts (Pelikán (2014)). Forecasting electricity production from hydro power plants tends to design more complex models to capture complex processes such as precipitation. On the other hand, the lower complexity of river stream flows leads to higher predictability as shown in Tongal and Berndtsson (2017).[132] Here, the complexity is measured on the data set as data (i.e. the time series) represent a sample of the phenomenon in question. Tongal and Berndtsson (2017) conclude determining the degree of complexity enables the pre-determination of a suitable model. It is implied artificial neural networks may be a better choice for high complex systems than deterministic models, especially for a one-step forecasting horizon.

Besides the natural complexity of processes behind forecasting objects, complexity increases with re-defining of the object; in demand forecasting in general, one may include returns of a product. However, this is not applicable in energy forecasting.

Water flow forecasting to optimize electricity generation

[132] Tongal, H. and Berndtsson, R. (2017). Impact of complexity on daily and multi-step forecasting of streamflow with chaotic, stochastic, and black-box models. *Stochastic Environmental Research and Risk Assessment*, 31(3):661–682

World weather models are also ranked in terms of their complexity degree, albeit without any clear definition of how to measure the degree of complexity, as in (Scher and Messori (2019)).[133] It is unclear whether a model with higher resolutions with fewer components could be assigned a higher complexity degree than a low resolution model with a larger number of processes.

[133] Their work defines complexity as a measure of the number of degrees of freedom of a system active locally around a given instantaneous state.

Regarding accuracy, combination of models and complexity, the relation between the increasing complexity of forecasting models and accuracy remains ambiguous. Akpinar and Yumusak (2016)[134] concluded that with the computation complexity accuracy rates increase. On the other hand, (Potočnik et al. (2014)[135] derived that "among the adaptive models, the nonlinear models did not surpass the performance of the adaptive linear models." Yu and Xu (2014)[136] addressed the issue of complexity by observing that "over the years, studies have shown that a combinative model gives better projected results compared to a single model for natural gas prediction" and saw the future of forecasting in hybrid models. Szoplik (2015) shared this vision. However, it is open to discussion whether combinations of models do increase their complexity in each case. Dimitriadou et al. (2018) concluded in their review of oil price forecasting that "machine learning methodologies produce a higher forecasting accuracy in comparison to the typical econometric ones and they typically outperform the random walk (RW) model,[137] while econometric approaches often

[134] Akpinar, M. and Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies*, 9(9):727

[135] Potočnik, P., Soldo, B., Šimunović, G., Šarić, T., Jeromen, A., and Govekar, E. (2014). Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia. *Applied Energy*, 129:94–103

[136] Yu, F. and Xu, X. (2014). A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Applied Energy*, 134:102–113. `https://doi.org/10.1016/j.apenergy.2014.07.104`

[137] Stock prices are often modelled as a random walk averaged over the number of stocks. By this, fluctuations are reduced.

fail to do so." Still, the latest M4 competition, organized by the International Institute of Forecasters in 2018 (Makridakis et al. (2020)), did not confirm Dimitriadou's conclusions.

Debnath and Mourshed (2018)[138] state "According to reviewed literature, NN (neural network) structure with two hidden layers produced best results for the monthly load forecasting, the peak load forecasting and the daily total load forecasting modules." However, there is no reasonably explainable connection between the number of hidden layers of the ANN network and the complexity of gas forecasting. Yalcinoz and Eminoglu (2005) state that 1) there is a relation between the *correct* number of hidden layers, domain (e.g. monthly load forecasting) and the accuracy, or 2) without any speculation of reasons behind it, until now two hidden layers produced the most accurate results for these domains.

[138] Debnath, K. B. and Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, 88:297–325. https://doi.org/10.1016/j.rser.2018.02.002

THE SHORT LIST OF SELF-CONTRADICTING STATEMENTS points out that without holistic consideration of a phenomenon to be forecasted, a data set as a sample of the reality and the complexity of the model, any statement is valid only for the study case used. The current state of the theoretical research is far from being able to generalize.

4
Study 1: German gas imports

Problem definition

The Science Direct search engine found 256 articles on gas demand forecasting (as of March 2020) and three articles on gas imports in relation to either China or India. Gas imports matter to foreign energy policy of any country with a strong industry sector (such as Germany or China), and for the strategy on pipeline/LNG[1] infrastructure development. This chapter underlines the features of gas imports in comparison to gas demand forecasts, and presents results of models (e.g. Holt-Winters filtering, regression analysis, and ANN) for monthly gas imports based on a self-constructed data set. Moreover, implications for energy engineering and IR are discussed.

[1] Interestingly, LNG gas is not a direct competitor to pipeline gas. Building the entire LNG chain results in further pipeline construction (Bridge and Bradshaw (2017)).

The International Energy Agency defines energy security as "the uninterrupted availability of energy sources at an affordable price" (International Energy Agency (2021))[2]; security of energy supply is a more specific term stressing the reliability of gas supply from production countries, routes, transit countries, and the infrastructure of an importing country. When imagining foreign policy as a four-dimensional idea (i.e. a country with time represented by stakeholders thinking and acting in the foreign policy field), the issue of energy security is one element competing with other issues in this space: combating terrorism, climate change, or pandemic diseases, political influence in supranational organizations, humanitarian help, military expenditures, etc. Thus, in countries such as Germany, with this space being overcrowded with a variety of issues, the topic of energy security is less prominent in IR than in countries synonymous with energy supply (e.g. Azerbaijan, Saudi Arabia).

[2] International Energy Agency (2021). Energy security. `https://www.iea.org/topics/energy-security` Last accessed 2021-01-09

In the last decade, various attempts have been made to quantify energy security by creating indices as in Biresselioglu et al. (2015).[3] In their work, supply security was quantified by the number of supplier

[3] Biresselioglu, M. E., Yelkenci, T., and Oz, I. O. (2015). Investigating the natural gas supply security: A new perspective. *Energy*, 80:168–176

countries, supplier fragility, and the subject of this research *the overall volume of imported gas*.

TO BE CLEAR ON TERMINOLOGY, the security of gas supply and the reliability of gas supply are not synonymous. The latter is measured in the average time (minutes per year) of the service not being provided for a customer; in Germany's case: 0.99 minute per year Bundesnetzagentur and Bundeskartellamt (2019). Back-up mechanisms keep the reliability of energy supply high, even if the security of energy supply is lower for a short period of a year.

THIS RESEARCH SHOWS TWO NOVEL APPROACHES to this topic: 1) aiming at imports instead of demand/consumption from the data science point of view; and 2) reflecting on the relevance of imports for the discipline of IR. The next subsections compute maximum German gas imports, provide reasons behind a lack of work on import forecasting, and discuss differences in assumptions for import and demand forecasting. After the description of data set construction, the results of selected models are presented. The last subsection gives concluding remarks on forecasting of gas, and on its link to energy security and further to IR.

Natural gas imports into Germany. Infrastructure considerations

In 2017, Germany was the *no*.1 ranked country in terms of gas imports worldwide, and was *no*.13 in the ranking of gas exporting countries (i.e. re-exports). There are three routes for gas imports from the Russian Federation:

- The Yamal-Europe system crossing Belarus and Poland, with a capacity of 33 bcm/y,
- Nord Stream, connecting Russia and Germany directly via the Baltic Sea, with a total capacity of 55 bcm/y,
- The Ukrainian gas transportation system with a total capacity over 100 bcm/y, crossing Slovakia, and the Czech Republic with the border node at Waidhaus, or through Slovakia and Austria (International Energy Agency (2020))[4].

[4] International Energy Agency (2020). Germany 2020. Energy Policy Review. `https://www.bmwi.de/Redaktion/DE/Downloads/G/germany-2020-energy-policy-review.pdf?__blob=publicationFile&v=4` Last accessed 2020-12-27

Figure 4.1 shows the excerpt of the Transmission Map 2019 with import, export and virtual nodes for Germany. To answer the question

"what maximum amount of gas can be imported into Germany per year within the technical limits of the infrastructure (border nodes)?"

this work calculates the upper limit by using maximum technical capacity in GWh/day per transmission pipe at the border provided

Figure 4.1: Transmission Capacity Map - Germany, ENTSOG (2019)

by transmission system operators (TSO). The transmission pipe capacities are defined to satisfy demand on a daily basis. It is assumed that the node can be operated 90% of time.[5] The largest import node (Greifswald) is an exception; whereas its technical capacity is at 1,570 GWh/d, usually only 618.8 GWh/d (capacity adjusted by the capacities of OPAL pipeline) is considered. The German grid cannot absorb more gas; any excess gas would have to flow to the Czech Republic and from there back to Germany (Federal Ministry for Economic Affairs and Energy (2019)). Still, construction capacities of pipelines are higher than reported technical capacities. Therefore, it would be plausible to use a factor 1 (fully utilized time of a year) while calculating the maximum technical capacity for a year. Another factor could be used to mirror the technical feasibility of the maximum imports

[5] This figure was chosen after re-calculating delivered gas to Germany from the Nord Stream pipeline and comparing it to the capacity of the pipeline. The gas market operates with thermal power units (GWh); the capacity of pipelines is defined in mass flows measured in billion cubic metres of transported gas per year. For the conversion, we used the heating value of the gas imported from the Russian Federation.

per node and a pipeline section by taking into account the capacity of compressor stations. As daily capacities for border nodes are reported at the maximum level, this factor has been disregarded.

IN THE SECOND STEP, this maximum technical physical capacity for imports to and exports from Germany has been compared with actual imports and exports published by BAFA[6] for the year 2018.[7] For 2018, based on the above-mentioned assumptions, Germany used ca. 44% of its technically maximum possible capacity on the import side (10, 045, 322 TJ); for exports it was 41.6%.[8] Reported exports from BAFA also include "Ringflüsse" (loops) - the amount of gas leaving Germany at one border crossing point (e.g. Olbernhau) and entering Germany at another point (Waidhaus).

[6] German Federal Office for Economic Affairs and Export Control

[7] Actual import depends on demand in two market zones for H-Gas (high-calorific gas) infrastructure imported from Russia and Norway and the L-Gas (low-calorific gas) infrastructure from the Netherlands.

[8] Real export in 2018 was 1, 563, 930 TJ (source: BAFA) and the technically possible calculated export based on the information from the Transmission Capacity map was 3, 758, 047 TJ.

TABLE 4.1 SHOWS PLANNED LNG-TERMINALS for Germany.

Name of the installation	Status	Start-Up Year	Nominal Annual Capacity (in billion cubic metre/year)	LNG storage capacity cubic metres LNG
Brunsbüttel LNG Terminal	planned	2022	8,00	240.000
LNG Stade GmbH	planned	-	5,00	-
Rostock transsphipment	planned	-	-	-
Wilhelmshaven	planned	2022	10,00	263.000

Table 4.1: LNG Import terminals in Germany. Source: Gas Infrastructure Europe (2021)

Missing forecasting on gas imports

MOST GAS FORECASTING IS CONDUCTED on an hourly-basis as in Su et al. (2019); studies on monthly gas consumption forecasting are scarce, and based on the literature review conducted for this research, papers on forecasting gas imports are non-existent. Therefore, the *next close* field of short-term gas forecasting is investigated with studies summarized in table 4.2. Whereas types of models and criteria remain the same, objects of studies vary in scope and in space, as does the type of the best performing model.

Abbreviations from the table 4.2: ISSA - Improved singular spectrum analysis, LSTM - long short-term memory, FARX - AutoRegressive model with exogenous variables

TABLE 4.3 COMPARES APPROACHES in forecasting gas demand and gas imports. The variable *gas prices* deserves attention; as gas prices used to be excluded from data sets as there was no expectation of an increase and some governments strongly regulated them for end customers.[9] Therefore, gas prices have a low information value for modelling a problem. The occurrence of events such as attacks on

In theory, the prices of all energy commodities decrease in the long run.

[9] This is still the case in China.

[10] In September 2019, two Saudi Arabian oil facilities were attacked, and their production - accounting for 5% of global production - was interrupted. As a result, Brent crude prices increased by 14.6% to \$ 69.02, and US crude oil by 14.7% to \$ 62.90 (Wearden (2019)). Ten days after the attack, one of facilities restored its oil production (Astakhova (2019)).

Author	Forecast subject	Models compared	Criterion	The best outcome
Wei et al. (2019)	Daily gas consumption London, Melbourne, Karditsa, Hong Kong	Back propagation neural network (BPNN), support vector regression (SVR), multiple linear regression (MLR) and other	MAE, RMSE, MAPE, mean absolute range normalized error (MARNE)	ISSA, LSTM
Chen (2018)	Day-ahead high-resolution gas demand and supply in Germany	Presented model: functional FARX compared to alternative autoregressive models	Relative MAPE, relative RMSE, for direction: mean correct prediction (MCP)	FARX -for average values of relative MAPE
Merkel G. D. et al. (2017)	Daily gas load for a utility in the USA 62 operating areas of local distribution companies in the USA 10 years of data	Linear regression, artificial neural networks, deep neural network	MAPE, RMSE	Deep neural networks
Akpinar and Yumusak (2016)	Gas consumption (commercial and residential consumers, city-level) in Sakarya, Turkey Daily data summarized as monthly January 2011- December 2014	Time series decomposition, Holt-Winters exponential smoothing, autoregressive integrated moving average (ARIMA), SARIMA	MAPE, R^2	ARIMA
Potočnik et al. (2014)	Day-ahead gas consumption, local distribution company in Croatia Data: consumption, weather data for 2 heating seasons 5 November 2011 – 26 April 2012 9 November 2012 – 31 March 2013	Benchmark models: random-walk, temperature correlation Linear models: regression method, auto-regressive models with exogenous inputs Non-linear models: neural network, support vector regression (SVR)	The mean absolute range normalized error (e), the adjusted R^2 measure	Support vector regression
Yu and Xu (2014)	Day-ahead gas consumption in Shanghai 15 November 2005 – 13 October 2008	Three-layer Back Propagation (BP) neural network structure, with a genetic algorithm	MAR, MAPE, RMSE	CCMGA–Im MBP mode CCM- cat chaotic mapping
Szoplik (2015)	Hourly peak offtake of gas in the following year for the city of Szczecin, Poland 1 January 2009 – 31 December 2011	ANN network (MultiLayer Perceptrons)	MAPE, RMSE nRMSE	ANN

Table 4.2: Short-term forecasting - case studies

Saudi Arabian oil facilities in 2019[10] and their impact on gas prices cannot be predicted by well-established forecasting tools of statistical learning. Moreover, gas markets are still affected by regional circumstances; the U.S gas prices may have reacted with a price decline due to higher production of oil with an associated gas, whereas the cross-border prices for Germany remained unaffected.

4.0.1 Data collection

First, Eurostat data was checked along with data issued by the BAFA. As for the Eurostat data, gas consumption per month equals inland deliveries as these are *calculated*. This means, the energy balance fits 100%, which is never the case due to, e.g. the unaccounted for gas, i.e. the gas quantity that remains after the balancing of all metered input and outputs calculated across the gas accounting period (e.g. a gas day).[11]

The BAFA has published the monthly amount of cross-bordered gas in terajoules (higher heating value) and its average price[12] since 1991. The price reflects the price in the import contracts; amounts of gas sold for a spot price are not included. The comparison of this data with energy consumption data from Eurostat reveals huge differences. The highest difference between these two values takes place outside the heating season (e.g. in July 2015, the cross-bordered imports were at 368, 687 TJ, whereas inland consumption was just 160, 509 TJ).

A CONSTRUCTED DATA SET covering monthly data from January 2002 to December 2018 refers to Germany with the following inputs: 1) heating degree days, 2) cooling degree days, 3) German imports, 4) cross-border price for gas 5) amount of gas used for electricity production, 6) production of gas within Germany, 7) German gas exports, and 8) gas storage balance. Gas imports are less affected by random factors such as extreme temperatures than gas demand. No data was found to be missing.[13]

THE MINIMUM SIZE OF THE DATA SET depends on the number of parameters to estimate for a statistical model and on the amount of randomness in the data (Hyndman and Kostenko Andrey V. (2007)). The latter can be described by the variance in a data measured, for example, by the range or quartiles. Similarly, Hastie et al. (2009) mention two criteria for the sufficient size of a training data set depending on the signal-to-noise ratio of the function to be used, and the complexity of the models used for fitting the data. In this work, as for the regression with seasonal dummies, $m + 2$ observations would be required when m denotes the number of months in a year (and subsequently

[11] For a research article on an application of online statistical control techniques for *unaccounted for gas* (UAG), see (Botev and Johnson (2020)).

Botev, L. and Johnson, P. (2020). Applications of statistical process control in the management of unaccounted for gas. *Journal of Natural Gas Science and Engineering*, 76:103194

[12] In German: Monatliche Entwicklung des Grenzübergangspreises

Inputs are described by definition, units and information source are included in Annex.

[13] There are several imputation methods for imputing missing data, the most simple one is to use the mean or median of non-missing values. However, any method introduces additional uncertainty into estimates and predictions. The additional uncertainty could be measured by doing multiple imputations for the creation of different training data sets (Hastie et al. (2009)).

	Gas demand forecasting	Gas import forecast
Choice of variables	1) Gas consumption (time series) 2)Variables such as population, industrial activity, and weather included. Typical variables: past consumption, temperature, days of week, month, seasonal information, wind data, GDP, holidays, humidity, the number of gas subscribers and gas price (Šebalj, Dario and Dujak Davor and Mesaric Josip (2017)).	1) gas imports (time series) 2) Excluding a variable "the amount of exported gas" if the domestic gas production is negligible with no optimistic outlook for domestic gas reserves.
Gas price	Price sensitivity in the industrial and electricity generation sector.	Cross-border prices follow long-term contracts. Impact factors: world price for oil and liquefied natural gas (LNG), weather, negotiations of new supply contracts in Europe, etc.
Weather as a variable	Importance of heating degree days showed in the seasonal component of the decomposition analysis. For the day-ahead gas demand forecasting in Denmark with a similar climate to Northern Germany, solar radiation was found an ineffective variable in terms of accuracy (Karabiber and Xydis (2020)).	Diminished importance For day-ahead forecasting, nominations instead of temperature shall be included as an input (Chen (2018)).
Market players	Gas demand relates to customer sections: residential, industry, commercial and energy use (e.g. electricity generation).	Gas traders and gas storage operators.
Infrastructure Market regulation influence	Dense infrastructure. For the long-term prediction, see the regularly updated Network Development Plan Germany. (Müller-Syring et al. (2013)).	Cross-border connections of high-pressure transmission pipelines, LNG regasification terminals.
Number of actors	In the residential sector: 31.09 mil. people (German-speaking population older than 14 years) live in a household with gas connection. The trend continues to rise (Pawlik V. (2019)).	There are 25 companies active as gas importers into Germany.
Technical specification	All pressure levels in the gas transmission and distribution network. No ceiling in terms of the maximum possible gas demand is set in gas forecasting; gas infrastructure is far from being fully utilized.	High-pressure gas pipelines. In the modeling part of the infrastructure (in contrast to the practice in forecasting), Eser et al. (2019) set ceilings for a given node to the maximum observed annual import at the node in the past.
Future research trends	Changing the time density – monthly predictions produced from daily forecasts (Akpinar and Yumusak (2016)). Case studies to prove the speculation – the accuracy of gas consumption forecasting is related to the gas consumption patterns of climate zones (Wei et al. (2019)).	Considerations about forecasting in the true forecasting phase although the control variables are unknown.

Table 4.3: Approaches to gas demand forecasting and gas imports' forecast. Study case: Germany

parameters), one parameter is needed for the time trend and $m+2$ is regarded as the theoretical minimum for the estimation (Hyndman and Kostenko Andrey V. (2007)). Therefore, 14 observations would be sufficient for a reasonable regression analysis if *there is no randomness in the data*. For Holt-Winters forecasting, the minimum size of the data set is $m+5$ observations; i.e. for this work, the minimum size would be 17 observations. For comparison, 204 observations are included in the data set.

Few authors worked with data sets including gas imports into Germany until now; for example Chen et al. (2018)[14] used high-resolution data for gas flows at three cross-border nodes. Gillessen et al. (2019) included a gas feed-in into GASOPT model, per grid node and per hour. To our knowledge, no other researchers have been using data related to gas imports in Germany for forecasting or modelling purposes. A data set used for this section includes *aggregate imports* through all cross-border nodes. Due to the domain (gas imports) and the aggregation, it is highly improbable that such a data set would contain zeros.[15] Still, this phenomenon is possible in intermittent (energy) demand or supply forecasting, as can be seen in the cases of electricity production from PV-panels or wind turbines, or the district heat demand of a single household, to name a few.

[14] Chen, Y., Chua, W. S., and Koch, T. (2018). Forecasting day-ahead high-resolution natural gas demand and supply in Germany. *Applied Energy*, 228:1091–1110

[15] Zero values in a data set result in infinite values for some accuracy measures (division by zero) or in undefined values due to the division zero by zero. Hyndman and Koehler (2006) provide circumvention to these problems by introducing a mean absolute scaled error (section Related Work).

For the short-term forecast *in a city*, gas prices, GDP per capita, and other economic factors do not impact the outcome (Szoplik (2015)). The gas price change causes a fuel-switch in the long-run only.[16] Still, a cross-border price remain included in the data set as it corresponds to the amount of imported gas to Germany. Both variables are derived from the same data source and contribute to data consistency.

[16] In the long-term; households may change their heating systems to heat pumps, solar panels, etc. due to 1) higher economic standards in the society or 2) a combination of policy measures for higher deployment of renewable energy and the society's acceptance to pay higher prices for environmental protection.

Domestic demand is the most important variable for imports; both the residential and industrial sector are of equivalent relevance in Germany. Demand in the residential sector depends on the outside temperature as gas is mostly used for heating (thus an input: heating degree days). An input "cooling degree days" is also included assuming that:

a) during hot summer days in cities; air conditioning in a commercial sector[17] increases electricity consumption as compression chillers are the most widespread cooling technology in Germany, and

b) to some extent, this consumption will be covered by electricity generated in gas power plants.

[17] In contrast with other industrialized countries, the share of air conditioning in households in Germany is less than 5% , in the commercial sector it is about 60%.

Gas exports have been removed from the data set due to the strong collinearity with gas imports.

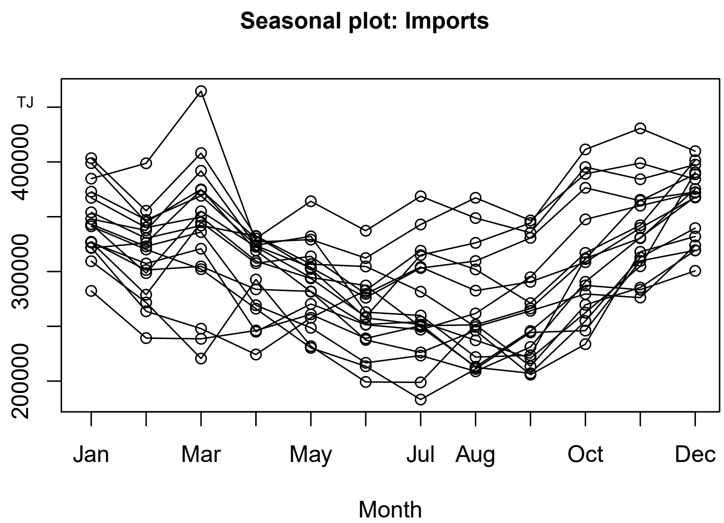


Figure 4.2: Seasonal plot of gas imports into Germany. The data set "Imports" covers physical flows of gas (TJ) in the period (January 2002 - December 2018).

IN THE SEASONAL PLOT, FIGURE 4.2, data are plotted against months to check how closely imports follow temperature differences. Some peculiar deviations from the expected line (based on the outside temperature) are due to the unequal length of months with the range from 28 to 31 days.

4.0.2 Material and methods

4.0.3 Time series models

For testing various models, out-of-sample accuracy measurement has been used by dividing data into a training set (January 2002 - December 2017) and a testing set (January 2018 - December 2018).

The time series "gas imports" was decomposed into trend, seasonal, and random (residual) parts (figure 4.3) with a distinct seasonal pattern. Furthermore, the data exhibits a rising trend in addition to monthly seasonality. For the sake of completeness, the data was also decomposed by using a multiplicative method, too – with results similar to those presented in figure 4.3.

In general, the objective of time series models is to discover the pattern in data series and produce forecasts by extrapolating that pattern into the future. Time series models are acceptable when knowledge of the prediction of the response variable is sufficient without the understanding of reasons for the change of the response variable.

WHILE CHECKING A TREND ON INPUTS, there is a seasonal variation

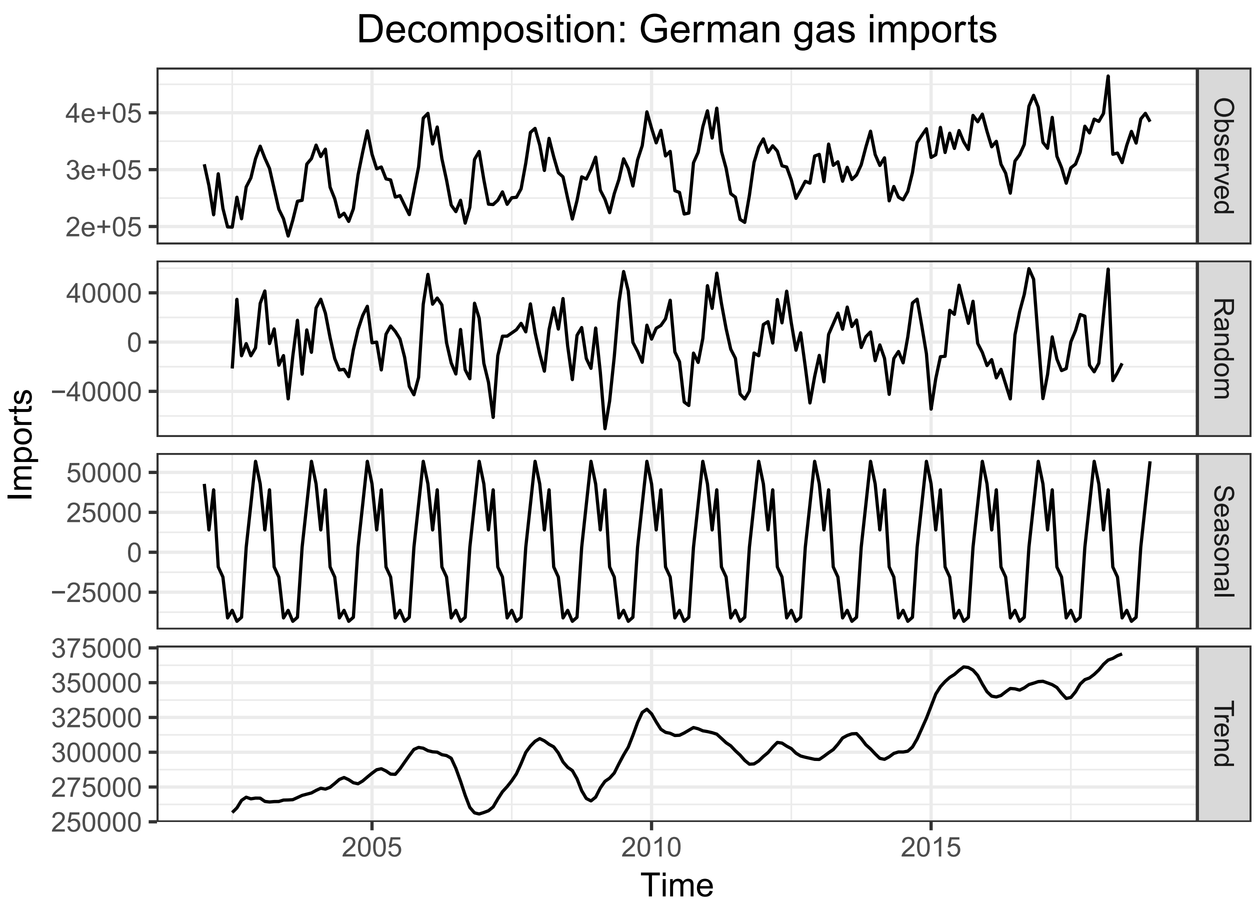


Figure 4.3: Decomposition of additive time series

in heating degree days, imports, gas for electricity production,[18] and storage. For the description of the time series, an additive model is appropriate as seasonal fluctuations are more or less constant in size over time. For "cooling degree days", the size of the seasonal fluctuations and random fluctuations seem to decrease with the level of time series. Hence an additive model would not be suitable. Second, the time series was seasonally adjusted by estimating and subtracting the seasonal component from the original time series.

[18] In the wholesale electricity trade, gas with its high marginal costs is next to the last or last option for electricity generation based on the merit order principle. Thus, this part of gas demand is sensitive to gas price changes.

As pointed out in the chapter *Related Work*, first a persistence forecast was computed, meaning a forecast equal to the most recent observation assuming a random walk model. A naïve forecast is optimal when data follows a random walk.

A naïve forecast is optimal when data follow a random walk (financial, economic data). Every "random walk" is unique and unpredictable but the ensemble of random walks possesses properties such as the central limit theorem. As steps N approach infinity, the endpoints of an ensemble of N step random walks with root-mean-square steps-size a has a normal probability distribution (Sethna, P. James (2020)): $\rho(x) = \frac{1}{\sqrt{2\pi\sigma}} exp(-x^2/2\sigma^2)$.

Any model needs to outperform the persistence forecast (figure 4.4) to justify the effort to construct it.

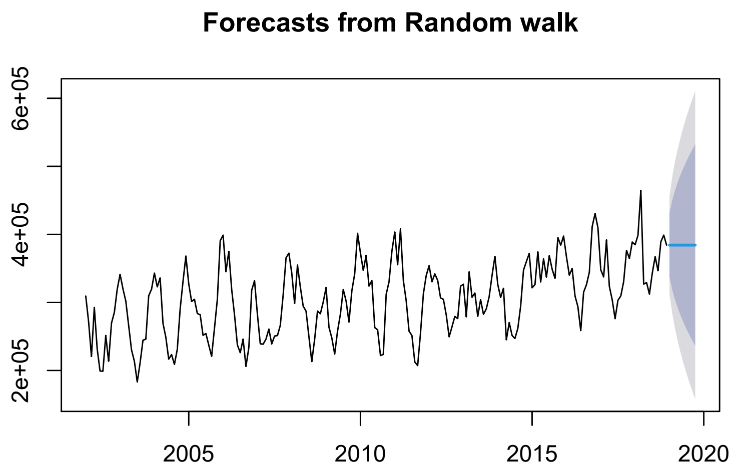


Figure 4.4: Persistence forecast. Widening intervals indicate increased uncertainty of future values (Fan chart).

Holt-Winters Time Series Model

The Holt-Winters method is the extension of the Holt method, capturing a linear trend, as it also captures seasonality in a time series by introducing a coefficient. The method applies exponential smoothing; smoothing equations for the level α, trend β, and seasonality γ are shown below. The seasonality can be modelled in an additive or multiplicative way:

HOLT-WINTER EQUATIONS FOR ADDITIVE SEASONALITY with the period length p[19]

$$\hat{Y}_{t+h} = a_t + b_t h + s_{[t-p+(h-1) mod p]+1} \tag{4.0.1}$$

where a_t denotes level, b_t denotes trend and s_t denotes seasonality. They are given by

$a_t = \alpha(Y_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1})$
$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1}$
$s_t = \gamma(Y_t - a_t) + (1-\gamma)s_{t-p}$

HOLT-WINTER EQUATIONS FOR MULTIPLICATIVE SEASONALITY; (level + trend) $\times$ seasonal component:

$$\hat{Y}_{t+h} = (a_t + hb_t) \times s_{[t-p+1+(h-1) mod p]} \tag{4.0.2}$$

where a_t denotes level, b_t denotes trend and s_t denotes seasonality.[20] They are given by:

[19] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10

[20] $\hat{Y}_{t+h}$ denotes the forecast for h periods ahead, p is the length of the seasonality, e.g. the number of months in a year.

$$a_t = \alpha(Y_t/s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t/a_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(Y_t/a_t) + (1-\gamma)s_{t-p}$$

SPECIAL CASES ARE OBTAINED while setting smoothing parameters α, β and γ to zero; with $\alpha = 0$ the level is unchanged, with $\beta = 0$ the slope is constant over time and with $\gamma = 0$ we fix the seasonal pattern. In the formula a_t, real values are deseasonalized by dividing real values by the seasonal number (Makridakis et al. (1998)). In the formula s_t for seasonality, Y_t denotes real values from the data set, thus containing seasonality and randomness, whereas a_t is already smoothed (average) value with a seasonality element. The factor γ serves to smooth the randomness which is necessary due to the Y_t. To start the algorithm, initial components must be set; for the HoltWinters function of the R *stats* package, start values come from a simple decomposition in trend and seasonal component using moving averages on the start.[21]

IN THIS STUDY, both trends showed similar results: the values for forecast imports were higher for the multiplicative trend but not more accurate. Parameters are determined by minimizing the squared prediction error and low values for β and γ[22] indicate that rather older values of x are weighted more. Values of α (specifying how to smooth the level component) near 1.0 mean that the latest value has more weight.

[21] R Documentation (2020). Holt-Winters function | R documentation. `https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/HoltWinters` Last accessed 2020-09-06

[22] For an additive model: α = 0.8293 (the coefficient for the level smoothing), β = 1e−04 (the coefficient for the trend smoothing) and γ = 1e−04 (the coefficient for the seasonal smoothing).

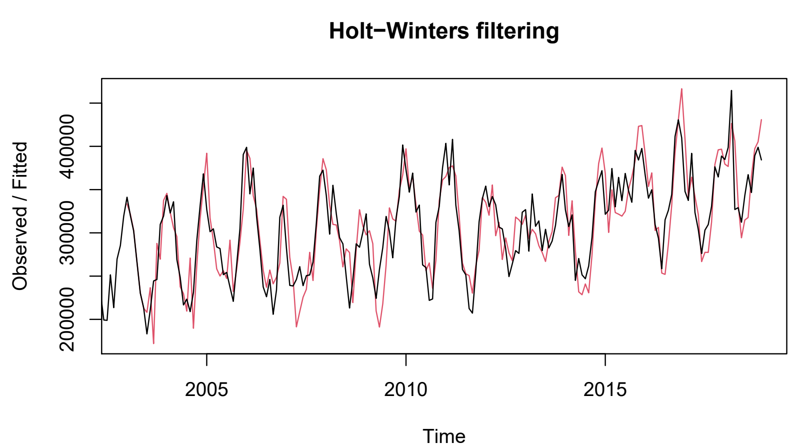


Figure 4.5: Holt-Winters filtering. The black line represents the actual value; the red line represents the filtered time series.

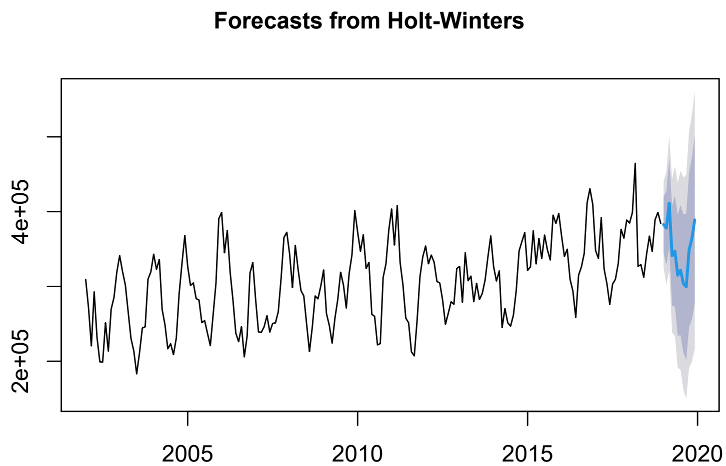


Figure 4.6: Point forecasts and the 80% and 95% prediction intervals obtained using Holt-Winters model

.

Figure 4.5 depicts forecasted values as well as the actual values. The method performs well in predicting seasonal peaks, i.e. the highest amounts of gas cross-bordered per year. The fan chart (figure 4.6) shows prediction intervals, set to 80% and 95% by default.

THE AUTOCORRELATION FUNCTION INDICATES patterns in in-sample forecast errors called residuals (figure 4.7).[23] As for figure 4.7: 1) the boundaries are set by definition, 2) monthly data are used; the time lag is expressed in years, 3) by default, the ACF at zero time lag is set to 1. The blue lines are set to 95% confidence interval (an estimate of a fixed but unknown parameter value). Correlation values outside of this threshold are *likely* to mean a correlation. "Likely" is stressed as "a priori approximately 1 out of every 20 correlations will be significant based on chance alone."(E. E. Holmes, M. D. Scheuerell and E. J. Ward (2020)).

[23] Autocorrelation function is one of the diagnostic checks for model uncertainty. For any suitable forecasting model, residuals left over after fitting the model should be white noise (Makridakis et al. (1998)).

Second, the Ljung-Box test results have been evaluated with the p-value 0.001293, which indicates the possibility of non-zero autocorrelation within the first 20 lags. To sum up the evaluation, as there is a pattern in the error residuals and due to the result of the hypothesis test, the model would not be the first choice for predictions.

Besides the Holt-Winters time series (HWTS) model, we considered the Bayesian Structure Time Series models, however, with less promising results.

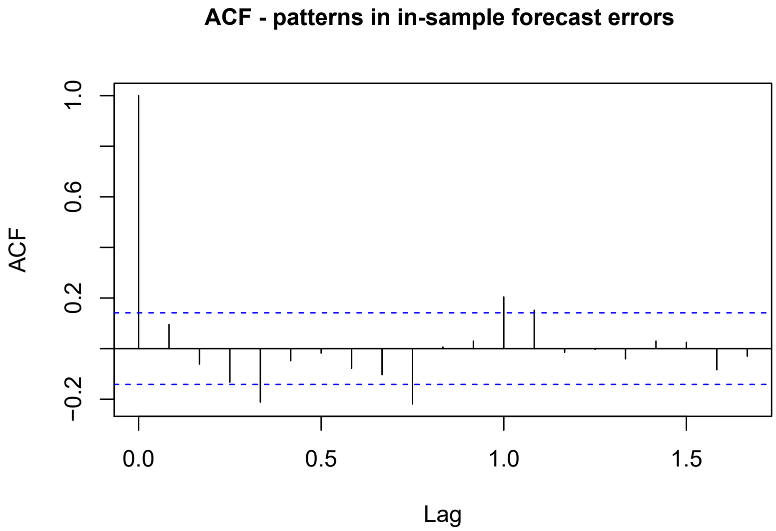
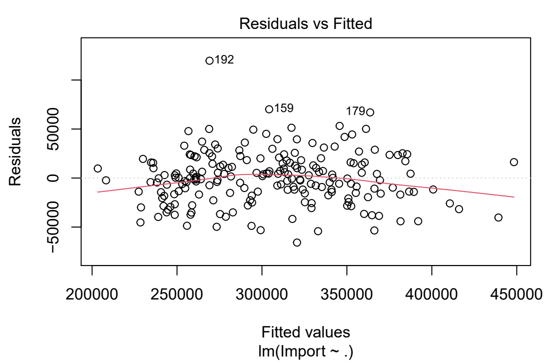


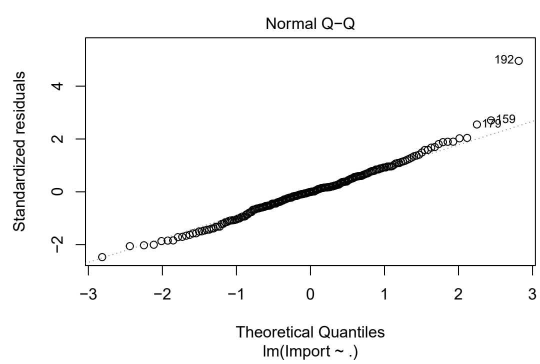
Figure 4.7: Autororrelation lag 1-20 (lag = 20)

4.0.4 *Regression analysis*

Regression analysis assumes that the output (gas imports) exhibits an explanatory relationship with a few independent variables under the assumption of continuity i.e. the explanatory relationship will not change. First, we included all variables from a data set, however, the variables *gas use for electricity production and cooling degree days* showed no significance in the model; hence they were excluded.

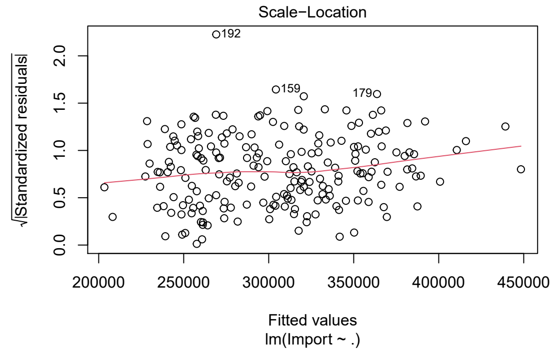


(a) Residuals vs fitted

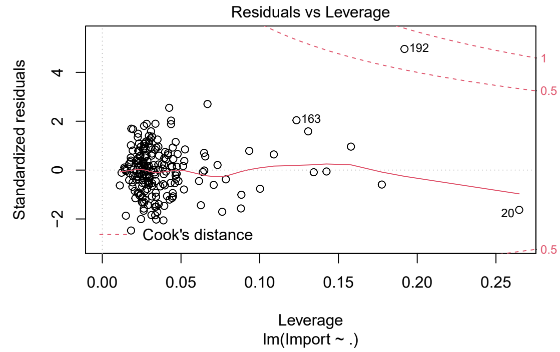


(b) Normal Q-Q

Three assumptions for the algorithm have been checked: the linear relationship, normality, and homogeneity. Residuals vs fitted depicted a horizontal line without a distinct pattern (figure 4.0.4, a); Normal Q-Q confirmed the normality assumption as residual points follow the



(c) Scale Location



(d) Residuals vs Leverage

straight dashed line (figure 4.0.4), b). Scale-location was used to check the homogeneity of variance of the residuals, depicting a heteroscedasticity[24] problem (figure 4.0.4, c).

[24] Heteroscedasticity refers to the situation in which the variability of a variable (residuals) is unequal across the range of values of a second variable (fitted values).

As for the feature selection,[25] analysis of variance (ANOVA test), has been conducted with two models; model 1 including all of the variables, model 2 with all variables but two above-mentioned inputs. Although feature selection has improved models, the standard residuals overall pattern change is negligible. To solve the problem of non-linearity, we also considered introducing an interaction between some predictors. In the Explanatory Data Analysis (EDA), a relationship between the consumption of gas for electricity production and gas storage balance has been detected. Therefore, an interaction between them has been introduced to determine a possible difference. However, this interaction, as well as the interaction of heating degree days and gas storage, has not contributed to the increased performance of a model.

[25] Feature selection is an optimization problem of searching for the features' combination that predict the response optimally (Kuhn and Johnson (2016)).

Also, dummy variables have not improved the model and eventually the standard regression models were chosen, excluding price.

4.0.5 Artificial Neural Networks

An artificial neural network computes its output by multiplying the inputs x by weights (w_0) and passing the result through an activation function. The training algorithm selects the weights of the input units by following the goal of minimizing a cost function, e.g. mean squared error (MSE). As the scaling of the inputs[26] determines the effective scaling of weights, data has been normalized. Within the R package neuralnet, the network architecture with the logistic activation function has been chosen under the criterion of the least total

[26] Usually all inputs are standardized to have a mean of zero and standard deviation of one (Hastie et al. (2009)).

squared error (figure 4.8). After the training step with the network learning an approximation of the relationship between inputs and an output (gas imports), predictions have been made and compared with the testing data.

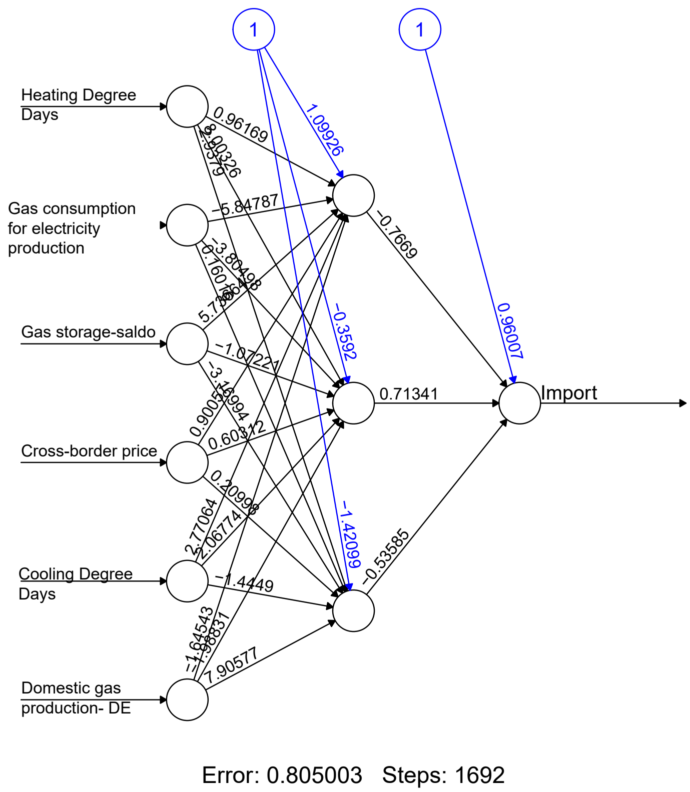


Figure 4.8: Plot of a trained neural network including synaptic weights and basic information. The black lines show connections between layers and weights on each connection; the blue lines show the bias term for each step. The training process needed 1692 steps until absolute partial derivatives of the error function were smaller than the default threshold (0.01). The information value of the plot is rather low; one cannot relate inter-steps of the model to making statements on the response variable.

A SMALL SIZE DATA SET IS NEVER IDEAL for ANN, as neural networks require a much larger number of observations than other models such as linear models.[27] Further, it is not clear which parameters (i.e. number of layers and number of nodes) are best for the network; in this case, one layer with ten nodes produced the best result in the mean absolute percentage error (MAPE).[28]

[27] Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and applications*. Wiley, Hoboken, NJ, 3. ed. edition. ISBN 9780471532330

[28] We also tested the random forest model. This approach uses the most recent data from the data set as the forecast for the next period but it showed over-fitting and it was left-out of consideration.

4.0.6 Modelling conclusions

The performance of models is checked by the accuracy measure MAPE as it is not scale-dependent, there are no zero values in the data set and the results are not intended to be compared with gas imports of countries with other pattern profiles. Table 4.4 provides an overview of the out-of-sample forecasting errors for three chosen models with the

Hyndman and Koehler (2006) recommends using MAPE if all data is positive and much greater than zero due to its simplicity.

persistence model where the forecast for all months of 2019 equals the value of the last observation from the data set, December 2018. Only results for the ANN with three nodes combat the persistence model and are comparable with results in Szoplik (2015).[29] No MAPE outcome is equal to or exceeds 50% of the Lewis Benchmark for inaccuracy (Lewis (1982)).[30] Lewis' benchmark is arbitrary; labelling the model as accurate depends on the object of forecasting, to be more general: on the model domain.

[29] Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208–220

[30] Lewis, C. (1982). *Industrial and business forecasting methods. A practical guide to exponential smoothing and curve fitting*. Butterworth, London. ISBN 0408005599

Table 4.4: MAPE comparison.

	Persistence	Regression	Time Series	ANN with 3 nodes
MAPE	0.0924	0.116	0.278	0.062

RESULTS ARE REFLECTED UPON IN THE RELATION TO STATEMENTS FROM PREVIOUS STUDIES:

- There is still a gap between the rapid pace of algorithms' development and their real-world application (Čeperić et al. (2017)).[31] Although algorithms work in many cases mentioned in the literature on testing the models (Merkel G. D. et al. (2017)), they are still rarely used as the basis for a decision-making process. As all results are valid for data sets (simplified sample of reality) with approximations describable by mathematical functions at the current stage of data science, it is hardly possible to make one step out of the boundaries of a model and make a statement about the observed reality.[32] Authors who dare to make this step, call their statements *speculations*. Increased complexity of emerging models stresses the relevance of interpretability.

- In real-time forecasting, all variables become the subject of uncertainty, and solving the issue by using rough estimates would produce relative confidence bounds so large that they would lose any practical use (Scarpa and Bianco (2017)).[33] Therefore, here we restrict a forecasting exercise to the testing data.

- Forecasting of gas imports is a more complex task than gas demand forecasting due to factors taking place outside the boundaries of the area (Germany). As an example, rising gas demand of neighbouring countries causes the steady growth of gas exports in the equation: *Natural gas imports = natural gas (internal) consumption + exports – natural gas own production +/- storages*; going hand in hand with the extension of infrastructure.

- An explanatory data analysis shows that the cross-border price does not influence German gas imports. In the short-term, the available

[31] Čeperić, E., Žiković, S., and Čeperić, V. (2017). Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy*, 140:893–900

[32] This is valid especially for neural networks and the field of medicine where the interpretation of models is crucial for the results' usability.

[33] Scarpa, F. and Bianco, V. (2017). Assessing the quality of natural gas consumption forecasting: An application to the Italian residential sector. *Energies*, 10(11):1879

gas infrastructure (i.e. technical physical capacity of entry points) sets the upper boundary for imported amount of gas.

- Yet, the gas market has all the attributes of a slowly-evolving mature sector: a long tradition with detailed legislation, regulation and standardization, a dense gas network, and sector independence. Still, the gas sector experiences pressure from both politics and the public to decrease its carbon dioxide and methane emissions and use its vast infrastructure system for transporting new products (hydrogen, biogas, synthetic methane). This could start a sort of decentralization of gas supply in the network: high pressure pipelines for the transport of the high-calorific gas (H-Gas) will be used for the rising export and distribution lines for local customers will be equipped with new supply entry points for synthetic methane and hydrogen injection. In some regions of Germany, the hydrogen infrastructure is being built as a project-specific island supply solution.

4.0.7 Geopolitical implications of gas imports

MONTHLY GAS IMPORTS matter for the security of energy supply. Thus, the issue is being discussed at state level.[34] The BAFA publishes total German gas imports on behalf of the BMWi (German Federal Ministry for Economic Affairs and Energy) and is not involved in forecasting. Several reasons behind *not* studying gas imports as a subject of short-term forecasting are:

[34] Most studies on energy security are country-specific (Ang et al. (2015)) although the concept of energy security requires a holistic view due to its interdependent character, especially in Europe.

- Gas network operators make internal plans for future consumption according to forecasts based on past consumption and weather forecasts.[35] It is unknown whether their algorithms are mainly based on ML methods.
- Ministries develop long-term strategies, where a one-year ahead forecast, based on monthly data, is too detailed for their goals (i.e. energy security).
- Germany's security of gas supply is high (BMWi (2019)). Due to data protection, the country of gas origin is not disclosed to the public anymore; however, this information is provided in the Bundesnetzagentur and Bundeskartellamt (2019) for 2017 data. Most of the gas has been imported from Russia, Norway, and – with a declining tendency – from the Netherlands.
- Missing high-resolution data on imports.

[35] Weather forecast providers use a deterministic model with no randomness, called an operational model and a probabilistic model running at specific hours and providing different weather scenarios, called an ensemble model (Gianfreda et al. (2020)).

In the IR discipline (academia), on the other hand, either unique events of the disruption of supply would be set into a story line or theories that enable understanding of main actors' actions are published.

In the previous work on gas supplies in the context of the IR discipline back in 2008, two assumptions were stated: gas supplies are concentrated and there will be no physical shortages in gas until 2020. Any shortfall of supply with demand will be due to political or economic obstacles.

Geopolitical debate on gas supply has undergone a shift in focus in the last twenty years. In the first decade of the 20^{th} century, the discussion in the IR circled around gaining control of new gas fields and transport routes. At Gazprom's Annual General Shareholder's Meeting on 30 June 2006, Alexey Miller, Chairman of the Gazprom Management Committee, stated:

Geopolitical perspective on supplies distinguishes a) pipeline gas with the geopolitical importance of transit states, b) LNG gas with no direct transit states but "choke points" such as Bosporus, Strait of Hormuz or Strait of Malacca.

"Ongoing global competition to gain control over hydrocarbon reserves has shown that state owned and backed companies have considerable advantages in obtaining dominating positions on international markets. Integration of state and commercial approaches enables to ensure a long-term planning based on prospective gas balance on the national and international scale." (Miller (2006)).

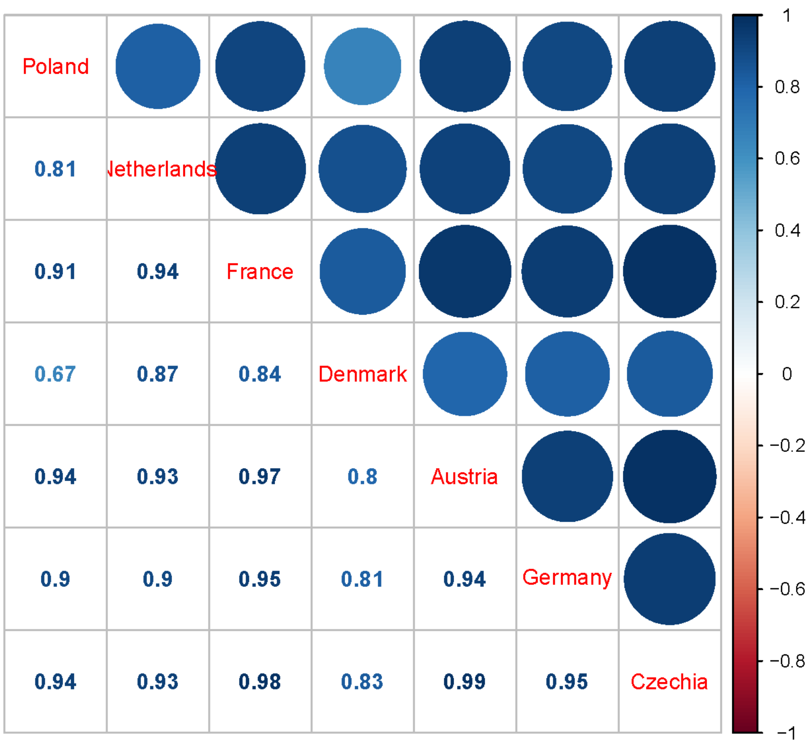


Figure 4.9: Correlation of gas consumption among Germany and its neighbouring countries (selection). Data source: Eurostat

The post-2020 global energy system still counts on rising gas imports but from new exporting regions (the US) and in another form (LNG). The intention for using gas has changed, too. With applying energy-efficiency measures within the residential sector, gas consumption will rise in the electricity sector to back-up the electricity production from renewable energy sources.

To sum up, high fixed costs for pipeline infrastructure and three separate gas consumption regions (Northern-American, Asian-Pacific and European) remained the reality of gas supplies. Changes occur in the gas pricing and the realization of producer countries that gas demand will be curtailed by pursuing environmental policies in Europe.

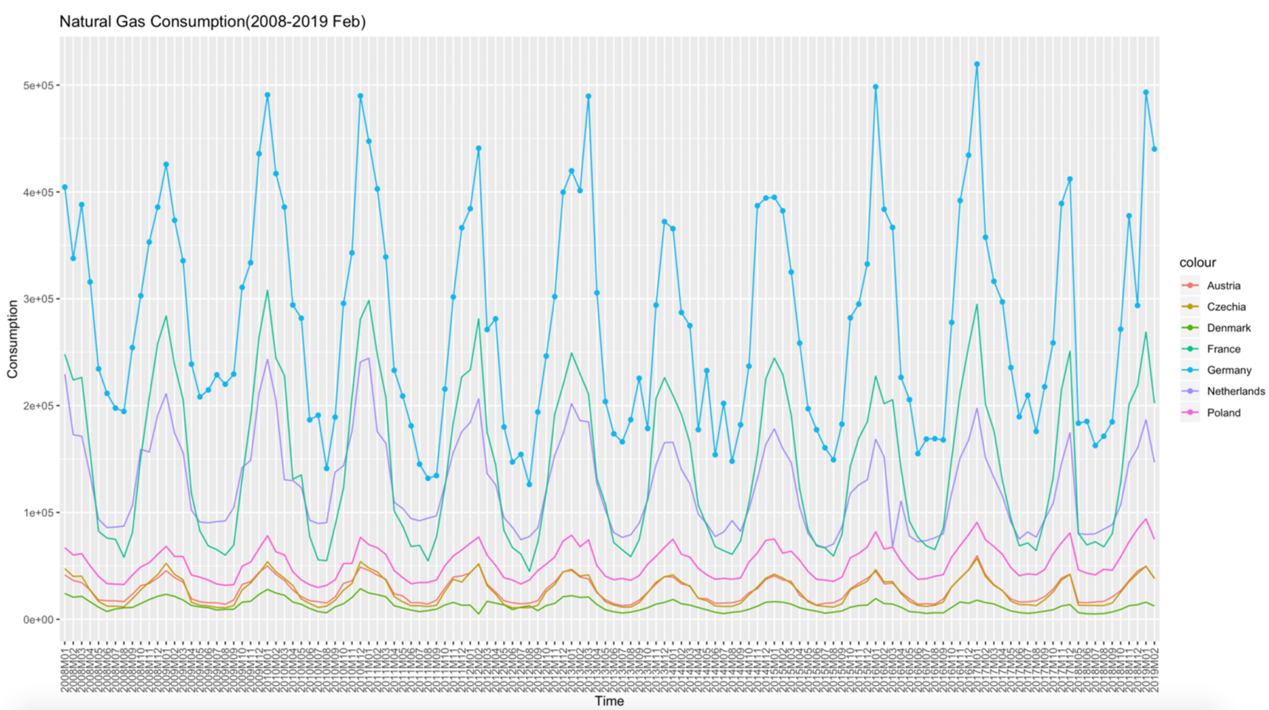


Figure 4.10: Time plot of gas consumption of Germany and neighbouring countries

Regarding Germany,[36] figure 4.9 shows the high correlation of gas consumption among Germany and some of its neighbouring countries. Shared climate and weather conditions are one of the reasons for the correlation. An analysis of of heating degree days revealed a high correlation among all countries in question but Denmark. Gas profiles in figure 4.10 mirror the set-up of the energy sector in a country; e.g. the curve for Germany depicts remarkable differences between winter peaks[37] and lows, whereas the Danish profile is flat as heating in Denmark is supplied by district heating based on renewable energy sources rather than by using gas boilers. Furthermore, gas consumption trends of France and Germany resemble each other. In France, industry consumes more gas than the residential sector as of 2016,

[36] Germany is an energy hub for *physical* gas flows in Europe. Gas consumption in neighbouring countries increases German gas imports while German gas exports also increase.

[37] Examples: extreme cold winter in 2010, a warm winter in 2014.

with the majority of gas consumed in the northern part of France.

GERMANY HAS BEEN MORE SUCCESSFUL in diversifying gas imports by routes than by product (diminishing L-gas imports from the Netherlands will be replaced by H-gas; low share of other methane-based gases).

5
Study 2: Yearly gas consumption in Germany

"The future is in a large measure a deconvolution of the past with seeds of novelty carefully selected and planted by the system itself." Cesare Marchetti

5.0.1 Introduction

In the second study case, this research constructs models such as principal component regression, support vector model, ANN in various variants for forecasting gas demand in Germany and tests how energy engineering knowledge impacts data set construction. The data set has been constructed assuming the link between the information gain from the energy sector of a country and the quality of energy forecasts. It is shown that data and the assumed relationships between a forecast and the explanatory variables must be country-specific in *energy forecasting.*

Biomass	Electricity from Biomass
Brent Price	Brent oil price
Coal Con	Coal consumption
Cool DG	Cooling degree days
Elect Gen	Electricity Generation
GDP	Gross domestic product
Heat DE	Heating degree days
Hydro E	Electricity from hydro power plants
Oil Con	Oil consumption
NG Con	Natural gas consumption
NG Imp	Natural gas imports
NG Price H	Natural gas price for households
NG Price In	Natural gas price for industry
NG_Prod	Natural gas production
NG_RES	Natural gas – proven reserves
Nuclear	Electricity from nuclear power plants
Pop	Population
Solar	Electricity from photovoltaics
Wind	Electricity from wind power plants

Table 5.1: Abbreviations. All variables relate to Germany, the Brent oil price excluded.

Defined boundaries

ANY EU COUNTRY HAS "the right to determine the conditions for exploiting its energy resources, its choice between different energy sources, and the general structure of its energy supply" (Article 194/2).[1] In Germany, oil and gas remain the only two sources affected by foreign events. The nuclear phase-out is scheduled to be completed by the end of 2022, and the coal phase-out by the 2030*s*. As a result, renewable energy as well as secondary energies will increase their share in the energy mix. Only gas fulfills the role of the bridge fuel in the vision of *Energiewende*, therefore oil has been excluded. Furthermore, the share of gas on primary energy consumption in Germany is substantial, amounting to 24% in 2018 and 25% in 2019. Thus, our primary focus lies on gas and Germany.

[1] EU (2012). Eur-lex-12012e: Consolidated version of the Treaty on the Functioning of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012E/TXT Last accessed 2019-07-22

5.0.2 *Forecasting literature*

MODELS AND METHODS ARE TREATED as synonyms in the literature,[2] and instead of differentiating between statistical/ML methods, this work follows Barker (2020).[3] She suggests a division between structured models (such as autoregressive models) and unstructured models (neural networks) based on how the knowledge is generated: is the process of generation defined a priori or is it learned from the data? Second, even studies from several regions cannot be compared with each other as the setup of every analysis with its forecasting models is unique in nearly all criteria, the time horizon being one of them. *Long-term forecasting* implies forecasting of technological change. The classical way of predicting technological changes is the Delphi technique, which is based on expert opinions. In the gas domain, experts would express their opinions on the rate of increase of hydrogen injection into gas pipelines provided that hydrogen has been produced in the water electrolysis process with electricity generated from renewable resources. The iterative process with the controlled feedback results in a consensus scenario which is perceived as the most likely development.[4]

[2] This is in contradiction to definitions in Fallah et al. (2019) (an article on the short-term load forecasting); a method is defined as an umbrella term for various models.

[3] Barker, J. (2020). Machine learning in m4: What makes a good unstructured model? *International Journal of Forecasting*, 36(1):150–155

[4] Dreborg, H. K. (2004). *Scenarios and Structural Uncertainty: Explorations in the Field of Sustainable Transport*. Trita-INFRA. ISBN 91-7323-068-5

STUDY 2 PRESENTS 1) the process of constructing a data set for forecasting gas demand, for Germany, 2) methods used for forecasting gas demand understandable within the scientific community of all backgrounds, and 3) demonstrates a first attempt to combine gained knowledge from both data science and social sciences.

5.0.3 *Research setup*

Constructing the data set and gathering variables

TYPICAL INPUTS FOR LONG-TERM GAS FORECASTING are GDP, (urban) population, energy consumption structure, energy efficiency or exports of goods and services (Zhang and Yang (2015)[5]; fuel[6], electricity prices as well as gas use for electricity production have been added to the list in this work. The complete data set contains 22 variables; they are described by definition, units and information source in Annex.

[5] Zhang, W. and Yang, J. (2015). Forecasting natural gas consumption in China by Bayesian Model Averaging. *Energy Reports*, 1:216–220. https://doi.org/10.1016/j.egyr.2015.11.001

[6] In the long-lasting scenarios in the Network Development Plan in Germany, assumptions for gas prices are divided into "international prices for oil and carbon dioxide" and "cross-border prices – Germany" (FNB-Konsultationsdokument), the latter was also used in the section on short-term forecasting of gas imports.

AS DATA WERE COLLECTED for a single purpose, they show some commonality in their origin. It is likely they occupy the same manifold space. This fact would gain importance when constructing an *unsupervised* neural network model where the manifold search space is not limited to a particular class, as is the case in supervised models. Second, the data set resembles a data set used in business rather than those used in forecasting competitions, as one can apply logical constraints such as all forecasts will be non-negative, use exogenous time series, etc. Third, the high quality of the data set with no missing values addresses partially one type of uncertainty:[7] the uncertain model input data. Still, as data in this research have been measured, summed up for Germany, and reported, they contain a systematic error.

[7] Three main types of uncertainty include parameter, model and data uncertainty. Regarding parameter uncertainty, a few models filter out irrelevant information. Interestingly, Petropoulos et al. (2018) tested six forecasting strategies for yearly, quarterly and monthly data of M and M3 forecasting competition for ETS and ARIMA models. Strategy 3, addressing all three types of uncertainties, failed to produce the most accurate results in terms of MASE and sMAPE in general and for yearly forecasts.

Petropoulos, F., Hyndman, R. J., and Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554

IN THIS RESEARCH SETUP, forecasting serves as a tool in business forecasting (technical infrastructure management). Pursuing forecasting methods forward would require standard data sets known from forecasting competitions (e.g. Kaggle, M4).

THE DATA COLLECTION PROCESS took place from December 2018 to June 2019. A few remarks about the variables:

- NATURAL GAS DEMAND, the response value. The energy balances of the federal states of Germany describe the total gas demand as primary energy consumption of natural gas, consisting of final energy consumption, the conversion input for natural gas power plants, losses and others. Figure 5.1 depicts the increasing variability of gas consumption (% change from previous year) in Germany, from about 2009 onwards.

- POPULATION represented as a total number of citizens, data was obtained from the International Monetary Fund (IMF), the World Eco-

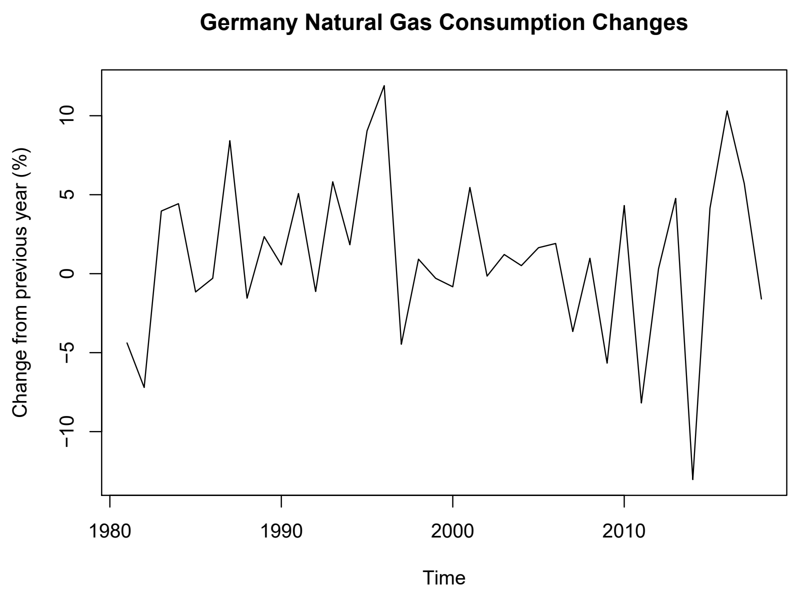


Figure 5.1: Increasing variability of gas consumption (per cent change from previous year)

nomic Outlook Database (WEO), and the Official Statistical Body (Statistiches Bundesamt). Data until 1990 refers to West Germany.

- Gross domestic product (constant prices), the same data source as for the population data. National accounts data until 1990 do not include FISIM (financial intermediate services indirectly measured). Data from 1991 refer to united Germany and include FISIM. Base year: 2010. The GDP deflator in the base year is not exactly equal to 100 since it is computed based on the quarterly real GDP data, which have been adjusted for seasonal and calendar effects. Type: Expenditure-based GDP, i.e. the total final expenditures at purchasers' prices, less the f.o.b. (free on board)[8] value of goods and services.

  In the simplified point of view, as the measure of net exports is included in real GDP (calculated by the expenditure approach), there is a hypothesis that energy security has a positive impact on the real GDP development (Šumskis and Giedraitis (2015)).[9]

- Gas prices, determined by long term contract prices and spot prices. The data set contains 1) average gas price for industry;[10] average price for households and total average price of gas, all without the value-added tax (VAT) but with the inclusion of usage charges and taxes, according to data from the Official Statistical Body in

[8] Free On Board (FOB) is a shipping industry term indicating the subject liable for goods that are damaged or destroyed during shipping.

[9] Šumskis, V. and Giedraitis, V. (2015). Economic implications of energy security in the short run. *Ekonomika*, 94(3):119. https://doi.org/10.15388/Ekon.2015.3.8791

[10] In Germany, industry consumes as much gas as the residential sector.

In German: Nutzungsentgelte

Germany (Statistisches Bundesamt).

- Prices of gas alternatives. The price for light heating oil in Germany has been added to the data set. Gas prices are not determined by oil prices any more,[11] still their corridor may be set with oil-prices being a ceiling and coal prices a floor. Villar and Joutz (2006) concluded the dynamics of the relationship as follows: a one-month temporary shock to the West Texas Intermediate (WTI) of 20% has a 5% impact on gas prices, but is reduced to 2% after two months. A permanent shock of 20% in the WTI leads to a 16% change in the Henry Hub price one year out *with all else equal*. In the model, a stable long-term relationship is assumed. Therefore, we included Average annual Brent crude oil prices from OPEC and the IEA.

In German: Preise. Erzeugerpreise gewerblicher Produkte (Inlandabsatz). Preise für leichtes Heizöl, Motorenbenzin und Dieselkraftstoff.

[11] Still, ca. 40% of gas sold in European markets is indexed to the oil price (Bridge and Bradshaw (2017)).

- Heating degree days, a factor for predicting consumed gas in the residential sector. Data come from Eurostat, until 1990, only heating days for West Germany are included. Cooling degree days in relation to gas consumption are especially relevant for countries with massive deployment of cooling inside buildings (the USA, Turkey etc.) due to peak electricity demand covered by gas engines. For checking the relationship of cooling degree days with gas consumption, this input has been included in the data set, too.

- Proven Natural gas reserves are defined as quantities that geological and engineering information indicates with reasonable certainty and can be recovered in the future from known reservoirs under existing economic and operating conditions. Data comes from BP; for the missing year 2018 data from the LBEG was used.[12]

[12] LBEG, Landesamt für Bergbau, Energie und Geologie, Erdöl- und Gasreserven in der Bundesrepublik Deutschland am 1. Januar 2019. On the 3rd July 2020 data sources' updates have been checked; a data value from the first source and the second one for 2018 is the same.

The L-gas production is re-calculated for energy content of 9.7692 kWh/cubic metre.

- Gas production. In 2018, gas production covered 6.5% of the gas consumption in Germany (Federal Ministry for Economic Affairs and Energy (2019)). The low calorific gas (L-gas) production in Germany as well as proven reserves decreases, with a decreasing trend predicted for both the production (table 5.2) and proven gas reserves. Gas production in Germany is one of few inputs for which future yearly values up to 2030 are available *with lower uncertainty*.

  As law in Germany does not permit shale gas production *on a commercial basis* and its costs are higher than the costs of purchasing pipeline gas now, shale gas production has been excluded.

  Data on gas consumption is taken from the BP Data Product "Energy Production and Consumption." As they are derived from tonnes

Year	Production Billion m^3	Capacity $Mil.m^3/h$
2019	6.26	0.80
2020	5.82	0.74
2021	5.72	0.73
2022	5.38	0.68
2023	5.11	0.65
2024	5.76	0.72
2025	5.44	0.68
2026	5.02	0.63
2027	4.61	0.57
2028	4.23	0.52
2029	3.99	0.49
2030	3.73	0.46

Table 5.2: Gas production prediction, source: FNB Gas (2019), original source: BVEG e.V (Bundesverband Erdgas, Erdöl und Geoenergie e. V.)

of oil equivalent (TOE) using an average conversion factor, they do not correspond exactly with gas volumes expressed in Germany's federal statistics.

- Data on natural gas imports is taken from the International Energy Agency (IEA) and from the Federal Office for Economic Affairs and Export Control (BAFA).[13]

[13] In German: Bundesamt für Wirtschaft and Ausfuhrkontrolle

- Electricity generation. Scenarios for Germany predict increasing share of *gas consumption for electricity production* due to the phase-out of nuclear and coal power plants. They balance electricity production from weather-dependent sources; to be precise, the lack of it. There are no data on gas consumption for electricity production back to 1980, therefore the data set includes electricity generation from BP. The value for the last year in the data set (2018) comes from the Federal Environment Agency (Umweltbundesamt). All data are based on the *gross* electricity production, i.e. power plants' own consumption is included.

- Other data such as oil consumption, hydro electricity, electricity produced in biomass power plants, by solar, wind and nuclear energy and coal consumption are from the BP Statistical Review of World Energy, 2019.

There is a lack of high-quality data sets on renewable energy sources, as they only started being collected in the 1990*s* (this research's data set starts in 1980). There is no data on LNG gas consumption; at the time of writing, Germany does not operate an LNG terminal. Still, even without an own terminal, the LNG availability does influence gas prices (and indirectly consumption) according to the survey among several traders and portfolio managers conducted in 2011 by Busse et al. (2012).[14,15]

After constructing the research data sets for a short-term[16] and a long-term period, relevant aspects are summarized in table 5.3.

5.0.4 *Methodology and Results*

Exploratory data analysis.

Data is highly correlated (the correlation plot in figure 5.2); thus, some of the correlated features have been removed. The data is not normally distributed and contains outliers. Since we keep any input due to the size of the data set, outliers have not been removed.[17]

For certain models, this research works with log transformations of skewed variables. Some models, such as support vector machines, do not assume the exact shape of distributions. Seasonality in the low frequency (yearly) data is not observed.

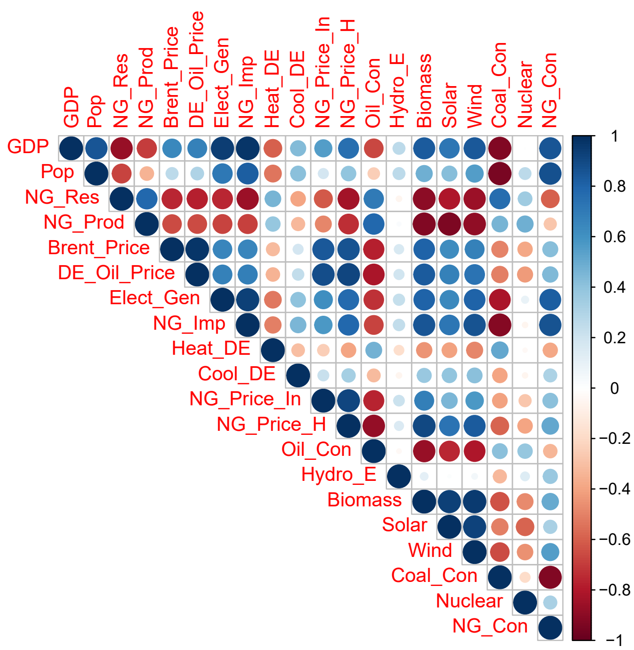


Figure 5.2: Variables correlation matrix

THE FOLLOWING FORECASTS WERE PRODUCED by multiple linear regression (full and reduced form), Principal Components Regression (PCR), Support Vector Machines (SVM), time series analysis, and neural networks. PCR addresses collinearity by reducing the dimensionality of the data set. The time series models applied poorly to the data

[14] Busse, S., Helmholz, P., and Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - an analysis of potential influence factors on basis of a NARX neural network. *Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012*. `https://publikationsserver.tu-braunschweig.de/servlets/MCRFileNodeServlet/dbbs_derivate_00027726/Beitrag299.pdf` Last accessed 2021-04-14

[15] Other factors influencing gas prices: geopolitical events and political decisions, seasonality/temperature, storage key figures, transport capacity, substitutes (e.g. oil price), spot and front month contracts for base- and peak-load) (Busse et al. (2012)).

[16] Previous section on gas imports

[17] Outliers could be defined as values differing more than four times the standard deviation from the mean as in Busse et al. (2012).

Criterion	**Aspect**	**Short-term forecasting horizon: hours, days, months, Frequency: high**	**Long-term with the yearly data Forecasting horizon: years, Frequency: low**
Purpose of forecasting	Use of forecasts	Improving operational efficiency of back-up plants for renewable energy sources, further saving energy and reducing emissions (methane and carbon dioxide)	Managing supply contracts, indigenous production (measuring gap between domestic production and consumption), infrastructure planning
Data set	Choice of variables	Weather variables (measured values or forecasts) – temperature or heating degree days, wind velocity, solar irradiance, air humidity Periods: weeks-days/public holidays	Macro-socio-economic indicators (population, gross national product), gas availability, gas and electricity price, price of competitive fuels/technologies.
Decomposition	Seasonal pattern	Yes, also within a day.	No seasonal trend in a classical statistical meaning.
Choice of the forecasting methods	Statistical methods vs judgemental methods	Extrapolation of past trends into the future (times series), Assumption: "patterns/relationships will remain during the forecasting phase.	Combination of statistical and judgemental methods; otherwise one conducts judgement only through method selection; extrapolation methods perform suboptimal due to missing seasonal trend. Periodic adaptation of a model by using adaptive models or by retraining the model from time to time.
Scenarios	Analysis of disruptions of supply	Possible if time steps involve hours, days and weeks	Insufficient for short-term disruptions analysis
Accuracy	Criteria	Mean absolute error (MAE), the mean absolute percentage error (MAPE), normalized mean square error (NMSE)	The same criteria as for the short-term or not defined. Scenarios are rarely evaluated ex post.

Table 5.3: Short-term and long-term forecasting of gas imports/consumption

and were abandoned. Neural networks were used in the full model, and two reduced models.

The training data set is composed of data from 1980 to 2011, and the validation data for the forecast ranges from 2012 to 2018, unless stated otherwise. The Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were computed and compared for each model.

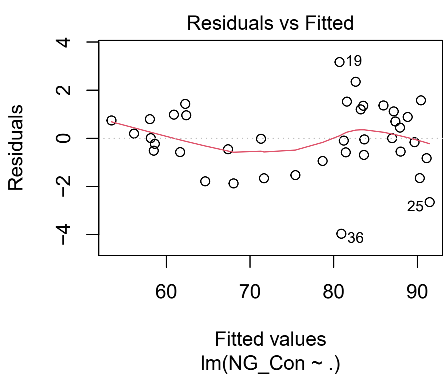
Linear regression

Full Model. The linear regression model should be tried first (Hribar et al. (2019)).[18] The first trial of a multiple linear regression with all variables shows the correlation coefficient R-squared adjusted at 0.9742 with most variables not significant and most coefficients close to zero. The variable Heating Degree Days has the highest significance level (p-value for the t-test) of 0.00421, followed by gas produced in Germany and electricity production from wind power plants.

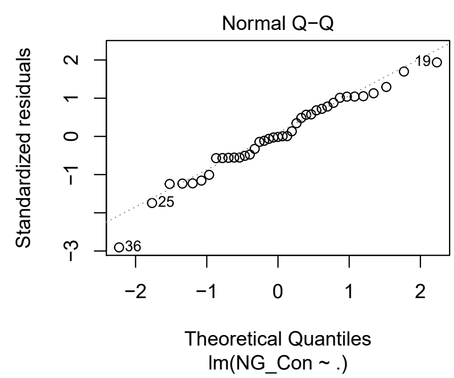
[18] Hribar, R., Potočnik, P., Šilc, J., and Papa, G. (2019). A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy*, 167:511–522

The high adjusted R-squared value suggests there might be an issue with collinearity or multicollinearity.[19] In the presence of multicollinearity, regression estimates become unstable, with higher standard errors.

[19] Collinearity implies two variables are near perfect linear combinations of one another, while multicollinearity involves more than two variables.



(a) Residuals vs Fitted plot



(b) Normal Q-Q

The residuals vs fitted plot appears (figure 5.0.4, a) to have a distinct nonlinear pattern in the scatter, suggesting that the relationship between the explanatory variables and the target is non-linear. The Q-Q (quantile-quantile) plot (5.0.4, right) is non-linear, indicating the assumption of normality is not met either. The full regression model performs poorly on the validation data (figure 5.3).

Reduced Model. The best subsets select which variables should be kept and which should be dropped from the linear model. A five-variable model with the minimum mean squared error prediction (MSEP) contains [NG_Prod], [Elect_Gen], [Heat_DE], [Wind], and [Coal_Con]. This model has an adjusted R-squared value of 0.9699 and all coefficients are statistically significant. Again, the research team checked diagnostic plots for testing the model's validity and concluded that model assumptions are not met. The analysis of variance (ANOVA test) of the full and reduced model shows the p-value is 0.249, indicating that the reduced model performs just as well as the full model.

Figure 5.4 suggests better performance although the p-value is large; for the ANOVA test, there are not large differences between the two models.

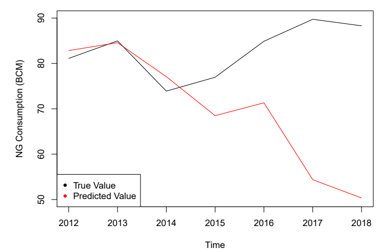


Figure 5.3: Prediction. Full linear regression model, 2012-2018

Removing variables correlated with the target. As the next step, variables [Coal_Con], [Pop], and [NG_Imp] were found to be highly correlated with gas consumption [NG_Con]. These variables were removed and the linear regression model was refit, resulting in an adjusted R-squared value of 0.9644 with few variables being significant.

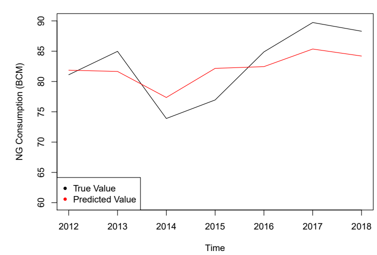


Figure 5.4: Prediction. Reduced linear regression model (2012-2018)

5.0.5 *Principal Components Regression*

PCR is one of the multivariate regresion models. "Multivariate" refers to more than one time-dependent variable.

Principal Component Analysis (PCA) converts a set of correlated variables into a set of linearly independent variables (principal components) using orthogonal transformations.[20] This technique is widely used when the number of variables exceeds the number of objects (in this research, years) or because of collinearities (Mevik and Cederkvist (2004)). To formulate the problem from the beginning; in multiple linear regression:

$$Y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + ...b_x \cdot x_x + \epsilon, \tag{5.0.1}$$

where x are predictor variables, y the prediction and ϵ error term. The β (beta coefficient, also called regression weight) is given by

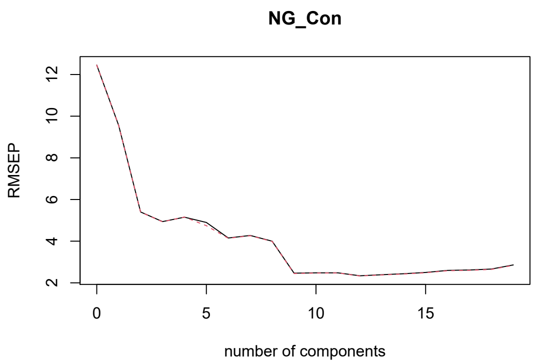
$$\beta = (X^T X)^{-1} X^T Y \tag{5.0.2}$$

Because of the above-mentioned reasons, the $X^T X$ could be singular. Such matrix is caused by linear interdependences among the variables i.e. some variable is a linear combination of other variables causing problems in a data analysis. Therefore we decompose X into orthogonal scores T and loadings P:

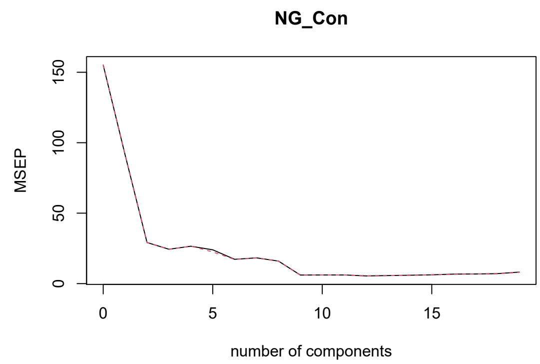
$$X = TP \tag{5.0.3}$$

[20] Each vector component is a linear combination of all variables and is orthogonal to other components in the set.

THE FIRST PRINCIPAL COMPONENT ACCOUNTS for the largest possible variance in data, and each following component has the next highest variance possible provided that it is orthogonal to the preceding components. Whereas in the standard linear regression model a dependent variable is regressed on the explanatory variables directly, the principal components of the explanatory variables function as regressors (independent variables).



(a) Cross-validated RMSEP curves for the gas consumption data



(b) Cross-validated MSEP curves for gas consumption data

SINCE PCR IS SENSITIVE TO VARIABLE SCALE, the data were scaled, i.e. standardised by dividing it by its standard deviation,[21] before finding the principal components. Using PCR, 91.22% of the variation in the data is accounted for by the first five principal components. The plots show curves of the cross-validated mean squared error (RMSEP) in figure 5.0.5 left and the root mean squared error (MSEP) in figure 5.0.5 right. If the Partial Least Squares Regression (PLSR), meaning the alternative, was used, few principal components would be sufficient for decreasing the RMSEP error. The PLSR reduces a dimension by not only summarizing the original predictors but also relating them to the outcome.

[21] The standard deviation is the square root of the average of the squared deviations from the mean.

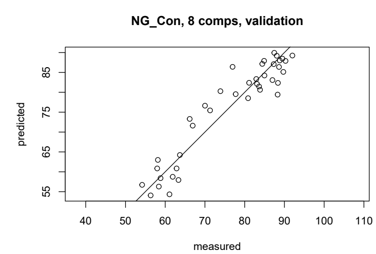


Figure 5.5: Cross-validated predictions for gas consumption data

INSPECTING A FEW ASPECTS OF THE FIT, with the first eight principal components chosen, the prediction plot in figure 5.5 shows cross-validated predictions versus measured values. The points follow the target line acceptably, significant anomalies have not beet spotted.

Built-in methods in the *pls* package enable use of a few methods for choosing the suitable number of components. As an example, one-sigma heuristic in figure 5.6 chose the model with the fewest components; one standard error away from the overall best model.

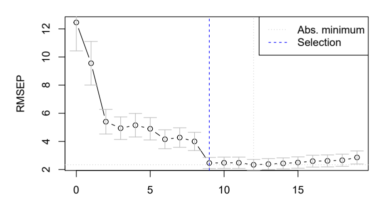


Figure 5.6: Example of one of strategy for finding optimal model dimension: one-sigma strategy

The first eight principal components accounted for ca. 98% of the data variation; results are shown in figure 5.7 depicting the overprediction of gas consumption.

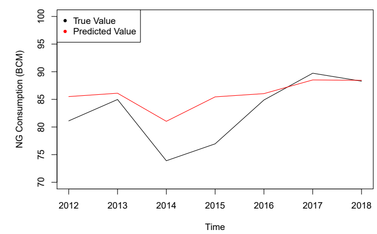


Figure 5.7: Gas Consumption forecast using PCR with the first 8 principal components (2012-2018)

To REVIEW, the primary value of forecasts lies in dealing with uncertainty while taking informative decisions. In the gas sector, decisions based on underpredictions may put the security of energy supply at risk, whereas decisions based on forecasts that turn out to be overpredictions decrease economic efficiency of the energy system. However, the energy infrastructure is usually built with a high security and safety margin.[22] Thus, while inventing a new accuracy measure for forecasting results for energy infrastructure, it would make sense to penalize the underprediction. By this procedure, forecasters would produce biased forecasts that could be accepted in the forecasting community. Keane and Runkle (1990) wrote on the asymmetric preferences over forecast errors:

[22] Examples: back-up power and heat plants, ability of the cold start for combined heat and power (CHP) plants. On electricity production company level, an overprediction can mean a government penalty and an underprediction an increased cost.

"If forecasters have differential costs of over- and under-prediction, it could be rational for them to produce biased forecasts. If we were to find that forecasts are biased, it could still be claimed that forecasters were rational if it could be shown that they had such differential costs."

Then, the assumption of symmetric loss function, present in standard regression-based tests would not be appropriate for evaluating forecasts with asymmetric preferences (Keane and Runkle (1990)).

5.0.6 *Support vector machines (SVM)*

CLASSIFICATION AND REGRESSION ANALYSIS are two fields for *supervised* SVM with no requirement for a normality assumption. We chose a non-linear ε -SVM with radial kernel function and tuning with the performance measure mean squared error. Kernel function separates two classes of data by transforming data into higher dimensional feature space which enables a linear separation (the kernel trick).

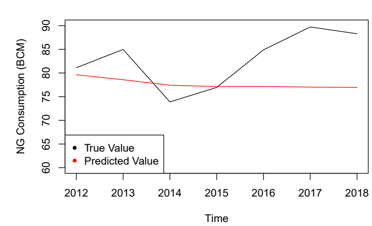


Figure 5.8: SVM full model, tuned (2012-2018)

The full model in figure 5.8 predicts almost the same consumption for each year. It underfits as the *best.tune function* chose a very low cost parameter C. This is in line with the literature on SVMs; "when the cost is small, the model will "stiffen" and become less likely to over-fit (but more likely to underfit) because the contribution of the squared parameters is proportionally large in the modified error function" (Kuhn and Johnson (2016)). Generally, this is the result of data not being scaled or the parameters not being tuned properly. As we already scaled data, this time we tuned parameters by extending the list of ranges for the cost parameter to 1000. Larger C leads to a smaller margin in the separating hyperplane, training errors will decrease and complexity increase. Obviously, decreasing training errors does not lead to decreasing the testing error (the distinction is explained in the section on complexity). The results of tuning are seen in figure 5.9.

Larger values of C address points near the decision boundary. A decision boundary is the region of a problem space in which the output label of a classifier is ambiguous (Hastie et al. (2009)).

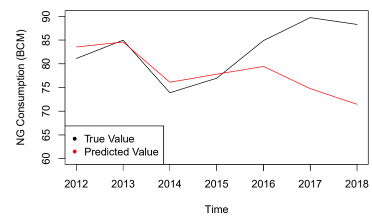


Figure 5.9: SVM full model, tuned II (2012-2018)

THE NUMBER OF VARIABLES IN THE SVM MODEL has been reduced by choosing variable importance greater than 50 ((NG_Imp, GDP, Coal_Con, Wind, Pop, Elect_Gen, Biomass) in figure 5.10.

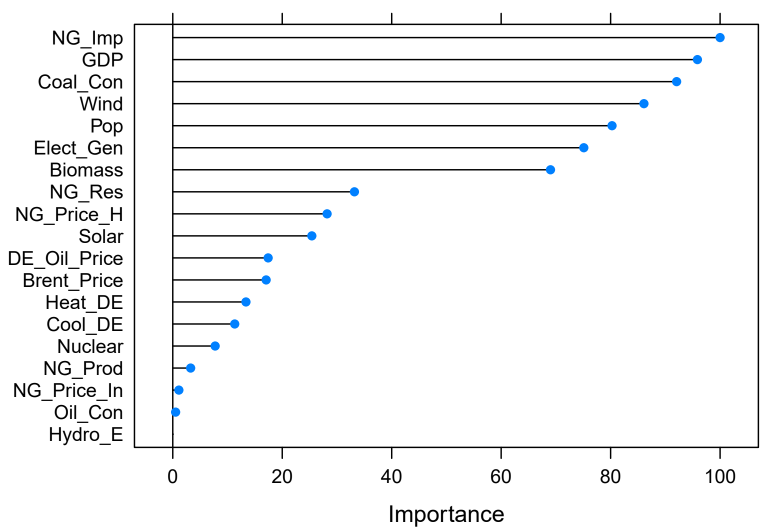


Figure 5.10: SVM Variable Importance

Variable importance indicates the most useful predictors for predicting the response variable *in this model context*. Though, the most important variables overall may not be the ones near the top of the plot. As figure 5.11 displays, overfitting prevents feature selection from producing more accurate predictions.

In the terminology used in Fallah et al. (2019), variable selection represents a method whereas the SVM is a model used for the method. Alternative models for variable selection include step-wise refinement, correlation-based methods (searching for variables correlated with the output but not with each other (Fallah et al. (2019)) and others. These methods partially rely on expert judgement.

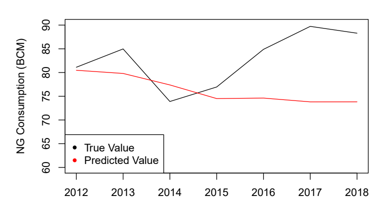


Figure 5.11: SVM reduced (2012-2018)

5.0.7 *Artificial neural networks*

ANN MODELS ARE UNSTRUCTURED models enabling forecasting of any subject without the knowledge of relationships between variables. However, some working packages have been built in the context of regression analysis and apply a *supervised* method, i.e. the known output is compared to the predicted input in the iterative cycles unless parameters of the model (called weights in ANN) are optimized, based on the minimization of the error function. Considering the size of the data set, and the fact that the goal is to forecast gas demand in Germany, the *neuralnet* package was chosen as it is understood as the extension of regression analysis (a supervised method).

THE FIRST STEP, DATA NORMALIZATION, enables the higher population of data for the same manifold space. Values for weights usually start randomly, drawn from a standard normal distribution (Günther and Fritsch (2010)[23]) and are adapted according to the chosen algorithm. The activation function transforms the node's aggregated input into the second layer. This works uses a non-linear logistic function $f(x) = \frac{1}{1+e^{-u}}$; the model was also run with the customized softplus function, approximating rectified linear unit (ReLU) function, however with less promising results. Two error functions could be used, either the sum of squared errors SSE[24] where $l = 1, ..., L$ indexes the observations, i.e. given input-output pairs, and $h = 1, ..., H$ the output nodes, or, if a classification problem exists, cross-entropy E (Günther and Fritsch (2010)).[25]

[23] Günther, F. and Fritsch, S. (2010). Neuralnet: Training of neural networks. *R Journal*, 2(1):30–39. https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf Last accessed 2020-12-20

An activation function is a differentiable function that is used for smoothing the result of the cross product of the covariate or neurons and the weights.

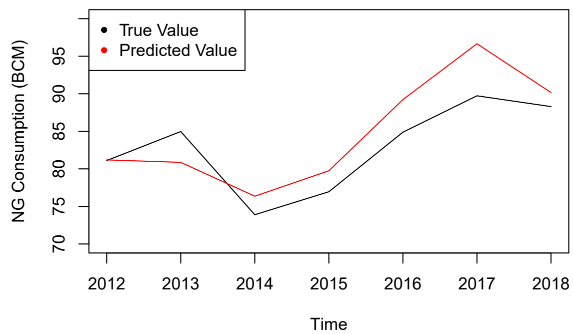
[24] $SSE = \frac{1}{2}\sum_{l=1}^{L}\sum_{h=1}^{H}(o_{lh} - y_{lh})^2$

[25] $E = -\sum_{l=1}^{L}\sum_{h=1}^{H}(y_{lh}log(o_{lh}) + (1 - y_{lh})log(1 - o_{lh}))$

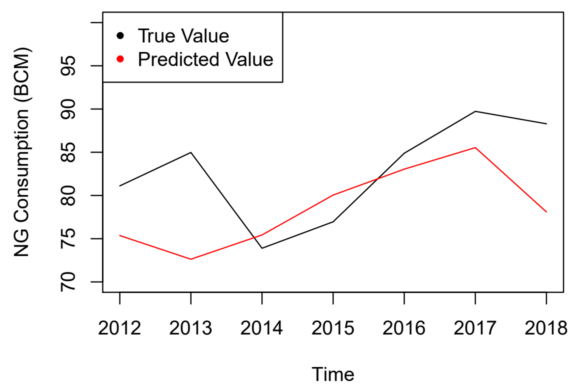
SEVERAL ALGORITHMS HAVE BEEN TESTED; the resilient backpropagation with weight backtracking (default) and without backtracking, as well as "sag" and "slr" inducing the usage of the modified globally convergent algorithm, authored by Anastasiadis et al. (2005). The latter further modifies a learning rate, either by associating it with the smallest absolute partial derivative ("sag") or the smallest learning rate ("slr").

"Sag" algorithm failed to predict gas consumption.

FIGURE 5.12 DEPICTS the forecasting using the full model with and without weight backtracking, figure 5.13 shows prediction with a classical backpropagation algorithm, used as a benchmark. Figure 5.14 shows results for the further modified algorithm by Anastasiadis et al. (2005) with the "smaller learning rate." The default choice (resilient backpropagation with weight backtracking) performs better than back-



(a) Prediction (2012-2018), ANN (the resilient backpropagation with weight backtracking)



(b) Prediction (2012-2018), ANN (the resilient backpropagation without weight backtracking)

Figure 5.12: The influence of weight backtracking on prediction

propagation without backtracking; the classical propagation and the method of Anastasiadis et al. (2005) does not follow the line of true consumption in these years.

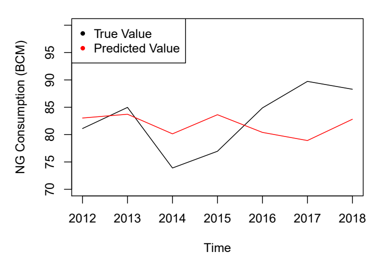


Figure 5.13: Prediction (2012-2018), ANN (backpropagation)

For any identified neural network, its weights follow a multivariate distribution and their confidence interval can be computed if the error function equals the negative log-likelihood (Günther and Fritsch (2010)).

An *identified* neural network does not include irrelevant neurons neither in the input layer nor in the hidden layers; i.e. variables with no effect on the response variable or variables that are a linear combinations of other variables shall be excluded. In the next subsection, this work presents several methods for models reducing the number of variables.

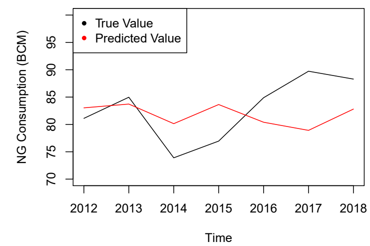


Figure 5.14: Prediction (2012-2018), ANN (Anastasiadis et al. (2005)),"Smaller learning rate"

Reduced models derived from the best case of the full model.

In the reduced model, two methods of variable importance select a subset of variables in the neural network. In the Reduced method 1, variables with an importance above 40 are selected (NG_Price_In, Nuclear, NG_Res, Biomass, GDP, Solar, Cool_DE, NG_Prod, Brent_-Price, Pop, and Hydro_E), figure 5.15. Gas consumption forecast using this method is shown in figure 5.16.

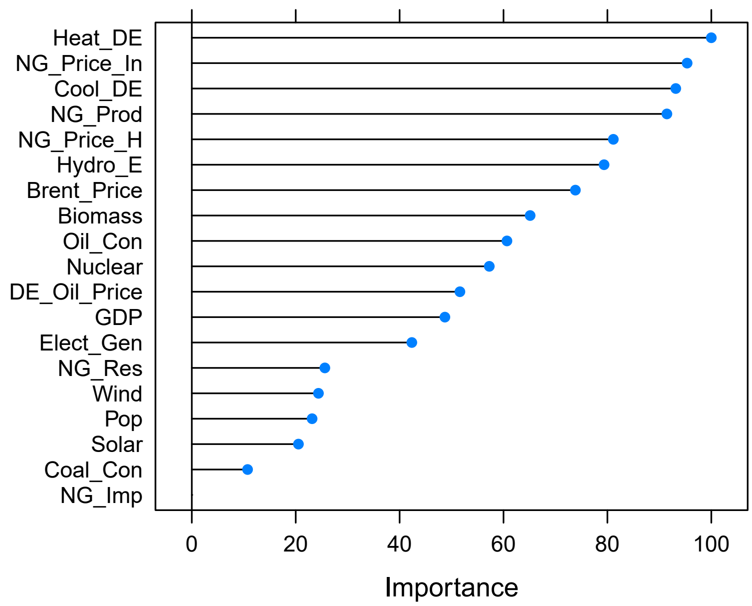


Figure 5.15: Neural network variable importance from method 1

In the Reduced model 2, variable importance was computed using one of the pre-defined set of functions: random forests of the rfe (recursive feature elimination) function of the package *caret*. The function uses the Algorithm 2 to prevent *selection bias* that could identify a non-relevant variable as important due to its random correlation with the outcome. That means, the genuine Algorithm 1 is placed inside of the layer of resampling, set by default to 10-fold cross validation.

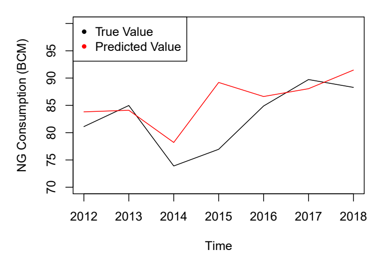


Figure 5.16: Gas Consumption forecast using ANN for the reduced model 1

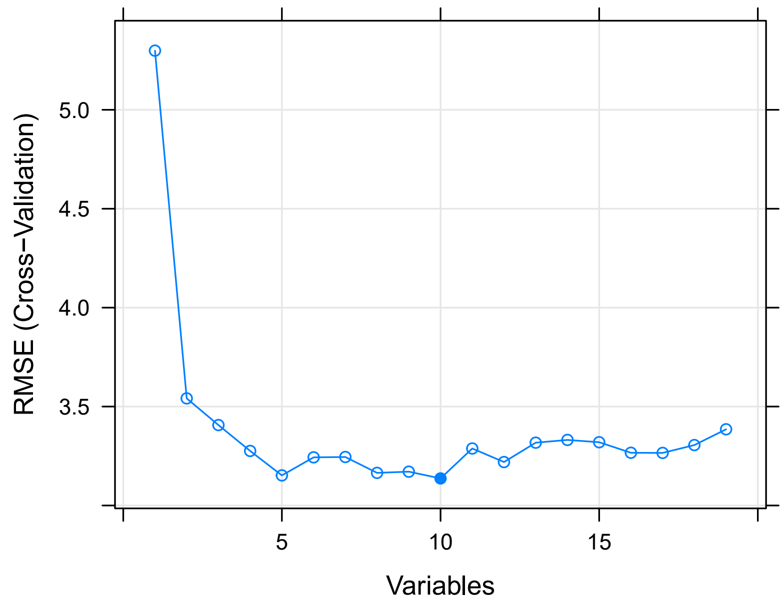


Figure 5.17: Neural network variable importance from method 2 (random forests), the performance profile across different subset sizes

The output shows that the best subset size was estimated to be 13 predictors (variables) as shown in figure 5.17: (Coal_Con, NG_Imp, GDP, Wind, Biomass, Solar, Oil_Con, NG_Price_In, Nuclear, Pop, Elect_Gen, NG_Price_H, NG_Res) and with three nodes in the hidden layer (results in figure 5.18).

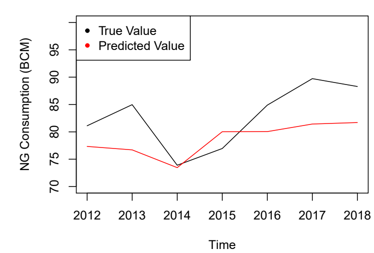


Figure 5.18: Gas Consumption forecast using ANN for the reduced model 2

It appears that the neural network and linear regression models perform best. However, the analysis also shows that the essential assumptions of linear regression are not met by the data; thus, the next section proceeds with neural networks.

THE PREVIOUS NEURAL NETWORKS work with the training data from 1980 – 2011, and predict seven years' worth of consumption values all at once. The next approach is to train the model using all previous data for *the given year of interest*. To predict for 2016, the model is trained using data from 1980 – 2015; however, to predict for 2017, the model is trained using data from 1980 – 2016. In the first approach we included only the data up to 2011 to train the model and the true value is not replaced with the predicted values for years 2012 – 2017 (figure 5.19), but in the second approach the additional *already forecast information* is used in the model. The second approach mirrors an actual forecasting method (ex ante - "before the event"), when the true consumption

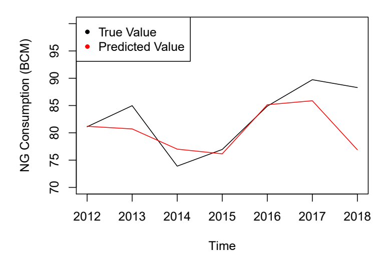


Figure 5.19: Gas Consumption forecast (2012 – 2018) using ANN for the full model, Approach 1

value for the previous year is unknown (figure 5.20). Ex post forecasts are produced from models that use the information available at time t other than the output (gas consumption) to forecast the gas consumption in the forecasting horizon $t+1$.

Although producing ex ante forecasts is rare in forecasting practice, they do matter in business and energy planning.

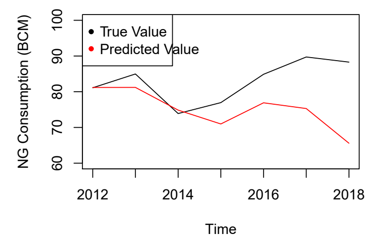
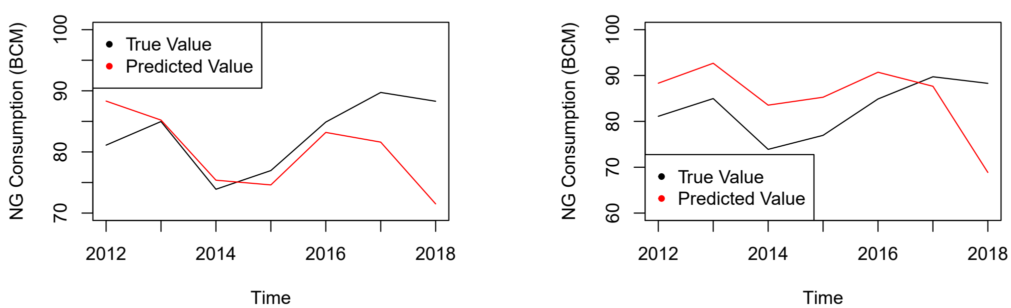


Figure 5.20: Gas Consumption forecast ($2012-2018$) using ANN for the full model, Approach 2

THE LAST EXPERIMENT PRODUCED RESULTS for one-year training, i.e. using only the data from the previous year (figure 5.21). That is, a full neural network is trained using the previous year data to predict consumption for the next year. This method performs well, with more inaccurate outcomes as the forecasting horizon lengthens.

As with the above strategy, to simulate how this would work in a forecasting scenario, the true values are replaced with the forecasted value for years 2012-2017. The results are in figure 5.21. Again, when using the predicted values to forecast, the results are less accurate. This is especially true for this method, since only the previous year is used to forecast for the next.



(a) Prediction ($2012-2018$), ANN, Only one-year data used, true values

(b) Prediction ($2012-2018$), ANN, Only one-year data used, forecasted values

Figure 5.21: Gas consumption forecast 2012-2018 (One year training with a) true values b) forecasted values)

Table 5.4 sums some relevant models and compares their MAPE and RMSE.

Model	MAPE	RMSE
Persistence model	0.6099	5.7427
Linear regression model, full	0.2549	20.5589
Linear regression model, reduced	**0.0411**	**3.6362**
Principal component regression	**0.0401**	**4.5759**
SVM, full model, tuned	0.0801	7.6122
SVM, tuned II	0.0826	8.8554
SVM, reduced variables	0.0999	9.3595
ANN, full model	**0.0372**	**3.8002**
ANN, reduced model 1	0.0441	5.3436
ANN, reduced model 2	0.0634	5.707
ANN, f. year by year with true values	0.0429	4.9759
ANN, f. year by year with predicted values	0.1123	10.9451
ANN, one year data, previous year, true d.	0.0700	7.6579
ANN, one year data, previous year, predicted d.	0.1068	9.9223

Table 5.4: Summary of values of evaluation criteria for main forecasting models created. Forecasts in the *Persistence model* equal the last observation from the training data set, Gas consumption in Germany in 2011. This "no-change" forecast provides the worst case upper bound on forecast error.

5.0.8 *Intersection gas supply security and forecasting*

THE CONCEPT OF ENERGY SUPPLY SECURITY is multi-faceted with clear geographic boundaries; the following example presents the idea in the gas supply context:

- GLOBAL LEVEL - Outside of the boundaris of energy supply security, the gas sector represents a major part of global energy policy for reducing global warming emissions. Reductions in the production, transport and consumption of gas contribute directly to the reductions of global warming emissions by reducing methane and carbon dioxide emissions. In this context, data scientists analyse data (time series) on failures of transmission and distribution pipes as well as the relationship between the rate of failures and the development of rules set by national legislation or by industry itself.[26]

- EU-LEVEL- Securing the supply of energy includes efforts to prevent disruptions, or coordination of states' actions when a disruption takes place. The revised Regulation no.2017/1938 (EU) defined the role for ENTSOG (The European Network of Transmission System Operators for Gas) to carry out union-wide simulations of gas

[26] As for national legislation: The German High-Pressure Gas pipeline Ordinance (GasHDrltgV), for industry rules: the DVGW Set of Rules as generally recognized Codes of Practice.

supply. ENTSOG simulates scenarios such as "Disruption of all imports to EU via Ukraine" and suggests measures on the supply and demand side (ENTSOG (2017)).[27] Their data sets include a winter period (October – March) with variables such as average daily demand and exports, average daily production, and storage (working gas volume and initial filling level).

[27] ENTSOG (2017). Security of Supply Simulation Report. https://www.entsog.eu/security-of-supply-simulation Last accessed 2021-04-14

- COUNTRY LEVEL- Strategic decisions on the construction of gas infrastructure, including LNG terminals, form part of a nation's energy policy. In the last decade, researchers have observed the growing importance of global market dynamics at the expense of geopolitics (Pulhan et al. (2020)).[28] Whereas in the past, researchers worked with notions such as scarcity of fossil fuels, oil supply peak and the goal of securing access to resources, current scenarios are based on the peak in demand for fossil fuels, probably taking place before the peak in supply (Nuttall and Manz (2008)).[29] Therefore, importing countries and supranational organizations shall consider revising and relaxing their geopolitical strategies. Countries that previously imported *only* are now seeking routes for exporting, the most prominent example being the US exporting liquefied natural gas. In Germany, although no LNG terminal is in operation now, several facilities are planned.

[28] Pulhan, A., Yorucu, V., and Sinan Evcan, N. (2020). Global energy market dynamics and natural gas development in the Eastern Mediterranean region. *Utilities Policy*, 64:101040

[29] Nuttall, W. J. and Manz, D. L. (2008). A new energy security paradigm for the twenty-first century. *Technological Forecasting and Social Change*, 75(8):1247–1259

- URBAN LEVEL - This is mainly discussed in academia. There are a few concepts of autarky for German cities such as Hamburg, Berlin, or Munich.

LINDLEY (2001)[30] IN THE PHILOSOPHY OF STATISTICS states "it is valuable to think about the relationships between the small world selected and the larger worlds that contain it.". *The small world* is defined by the data set size, chosen variables, and forecasting models. It recalls a mental picture of moving in one direction on a 20-cm-wide path with no option of widening this path for ever-growing feet (i.e. ever increasing complexity in a world or our increasing understanding of this complexity) but with a meticulous survey of anything found on this path. This process is supposed to enable the prediction of next steps. This is the best known way currently and it produces forecasts. *The larger world* is composed of the energy sector of a country, the economy of a country and the world economy.

[30] Lindley, D. V. (2001). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337. https://doi.org/10.1111/1467-9884.00238

The term used firstly in the meteorological context by Lorenz. To cite Robert M. May: "Alternatively, it may be observed that in the chaotic regime arbitrarily close initial conditions can lead to trajectories which, after a sufficiently long time, diverge widely. This means that, even if we have a simple model in which all the parameters are determined exactly, long term prediction is nevertheless impossible." (May (1976)).

SUPPOSE DATA IS AVAILABLE FOR ALL SITUATIONS related to one subject to forecast, such as the population of a single country, electricity load, or future conflict. This starting statement is already problematic as *related to* could be expressed only as a spectrum and even the smallest “amount” of relation would still count bringing a change in the outcome, as observed in the butterfly effect phenomenon. Thus, boundaries on *what to include* are blurred. Beside other factors, the subject to forecast (i.e. energy demand for a respective country) depends on the world economy, migration, changing lifestyles, and technological change, to name a few. The speed of the two latter factors is country-specific, as well as the final type of use of a commodity. If relations among them are incomprehensible for including them into one single model, an expert judgement remains the only option. Still, "models" behind expert judgement are a black box, again and as Smil (2000)[31] proved it - expert judgement inclines to linear thinking. If two to five observations are available, experts tend to see patterns in the data (Goodwin (2020)) expecting current world trends to continue rather than any rapid disruptions.

[31] Smil, V. (2000). Perils of long-range energy forecasting. *Technological Forecasting and Social Change*, 65(3):251–264

5.1 Implications for the energy sector

5.1.1 Hydrogen

There are multiple reasons for increasing hydrogen's share to cover energy demand in Germany and worldwide:

- to replace fossil fuels with synthetic fuel produced in power electrolysers that split water into its components, powered by electricity produced by renewable or low carbon dioxide emission sources. This would avoid costs in electricity infrastructure in the long term and stabilize electricity grids in the short term as hydrogen is produced when the spot price of electricity is low,[32] which indicates a higher share of electricity produced by renewable energy (Grüger et al. (2019).[33]
- to increase self-sufficiency of the energy sector.

Most hydrogen in Germany is produced from natural gas, only 7% (3.85*TWh*) of demand is met via electrolysis (Ministry for Economic Affairs and Energy (2020).

[32] In the simulation of Grüger et al. (2019) hydrogen production costs were reduced by up to 9.2 per cent and wind energy utilization increased by up to 19 per cent.

[33] Grüger, F., Hoch, O., Hartmann, J., Robinius, M., and Stolten, D. (2019). Optimized electrolyzer operation: Employing forecasts of wind energy availability, hydrogen demand, and electricity prices. *International Journal of Hydrogen Energy*, 44(9):4387–4397

TECHNOLOGIES FOR UTILIZING HYDROGEN have been developed for decades; the issue of right timing or maturing the idea have been assumed as reasons for their late deployment in the energy system. Breaking down segments of gas and hydrogen consumption brings to light issues with their mutual interchangeability. The highest segment in consuming gas in Germany is industry (30.6%), followed by the residential sector (28.8%), heat and power generation (24.4%), commercial (13.8%), other energy (1.8%) and transport (0.6%).[34] The total hydrogen consumption amounts to ca. 55 TWh with the highest share of demand in material production processes in industry and production of chemicals (Ministry for Economic Affairs and Energy (2020)).

[34] All shares refer to 2017. Source: International Energy Agency (2020).

Estimations of hydrogen use in the energy sector

BEFORE 2030, none of the German scenarios in Jensterle et al. (2019) assume extensive use of hydrogen. Some scenarios expect hydrogen demand at 133 PJ (36.94 TWh) in transport and industry. As for 2050, German scenarios count on the demand ranging from 300 to 600 PJ (83.3 to 166.7 TWh).

BY 2030, another estimation counts on increasing hydrogen demand by 10 TWh (Ministry for Economic Affairs and Energy (2020)) which does not automatically decrease the amount of natural gas used. Furthermore, planned increase of hydrogen in the transport sector does not impact overall gas consumption.

UNTIL NOW, natural gas has been considered to be transported via pipeline infrastructure in Germany and the whole gas consumption has been related to it. To decrease the carbon footprint of gas transported via pipelines and, thus, save the investment in infrastructure, hydrogen is considered as the next energy carrier to fulfill *energy demand*.[35] The first option discussed is injecting hydrogen into the pipeline system, bearing in mind limits in concentration due to technical constraints.

Altfeld and Pinchbeck (2013)[36] see a mixture up to 10 vol.% as not critical *in most cases.* In terms of combustion parameters, 10 vol.% of hydrogen decreases the Wobbe Index[37] by 3%, as it also decreases the methane number[38] and increases turbulent flame speed. Underground storage (bacterial growth), CNG steel tanks (interaction between hydrogen and steel), gas engines (increased combustion and end-gas temperature - higher NOx emissions) and gas turbines (no extra margin for variation in a fuel's properties) have been identified as sensitive components of the gas system in relation to higher blending of hydrogen with gas (Altfeld and Pinchbeck (2013)).

[35] Hydrogen will be utilized in all energy sectors: electricity production, transport, heating and industry.

The second option is to transport hydrogen through newly-built hydrogen pipelines.

[36] Altfeld, K. and Pinchbeck, D. (2013). Admissible hydrogen concentrations in natural gas systems. *Gas for Energy*, (03). https://www.gerg.eu/wp-content/uploads/2019/10/HIPS_Final-Report.pdf Last accessed 2020-12-19

[37] The Wobbe Index is the measure for the interchangeability of different fuel gasses.

[38] The methane number describes the knock behaviour of fuel gas in internal combustion engines.

5.1.2 *Publicly available estimations of gas consumption in Germany*

NATURAL GAS CONSUMPTION *at the state level* is reported either in billion cubic metres or exo-joules (both in the BP (2020)[39]) or in the energy unit TWh (Nymoen Strategieberatung (2016)[40]. For the conversion for past data, one needs to access the share of L-Gas with the heating value at 8.9 kWh/cubic metre and of H-Gas with the heating value of 11.1 kWh/cubic metre. Based on available information on the domestic production in Germany (L-Gas) and imports from the Netherlands (L-Gas), Norway (H-Gas) and other countries (mostly H-Gas) in 2016, shares have been calculated at 28% for L-Gas and 72% for H-Gas. The difference in 3.7% points rather to different methodologies used in energy statistics. As for 2030, all gas used will be H-Gas.

NYMOEN STRATEGIEBERATUNG (2016) DEVELOPED THE SCENARIO for final energy consumption until 2035 based on the rise of the GDP, population, forecasts for energy carriers, policy measures as well as on the forecasts of carbon dioxide emission certificates. Figure 5.22 shows scenarios for gas consumption development. Gas consumption (in primary energy) was supposed to decrease by 38 TWh, or 5%, from 2015 to 2020 due to climate protection measures (Nymoen Strategieberatung (2016) which the latest gas statistics do not confirm.

[39] BP (2020). BP Statistical Review of World Energy. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-natural-gas.pdf Last accessed 2020-12-21

[40] Nymoen Strategieberatung (2016). Strategische Marktprognose Erdgas. https://zukunft.erdgas.info/fileadmin/public/PDF/Politischer_Rahmen/marktprognose-erdgas-2016.pdf Last accessed 2019-12-27

The IEA reports typical gross calorific (higher) heating values of natural gas in MJ/cubic metres. Russia: 38.23 MJ, Norway 39.24 MJ, UK 39.71.The heating value of hydrogen is 12.7 MJ per cubic metre).

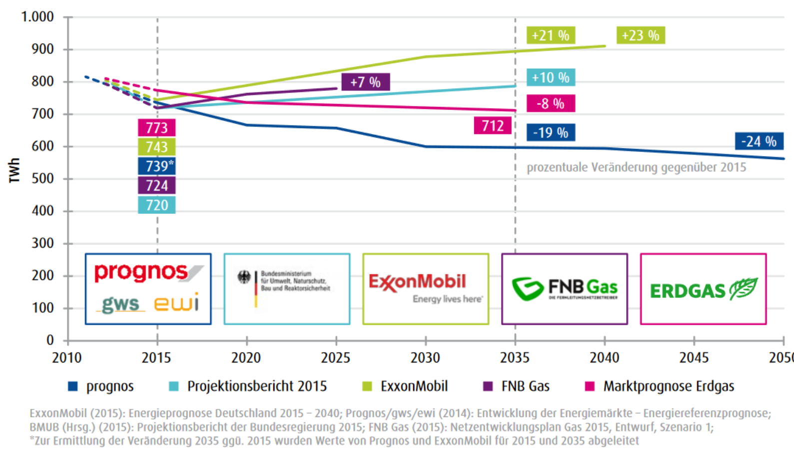


Figure 5.22: Market prognosis for gas consumption. Source: Nymoen Strategieberatung (2016)

Federal Ministry for Economic Affairs and Energy (2019) expects a slight to medium-sized decline in gas consumption by 2029, based on the Network Development Plans for Gas Infrastructure. Some German regions will experience an increase in gas consumption; therefore gas infrastructure will continue to expand and became more complex due to injection of synthetic methane, bio-methane and hydrogen.

Considering the ambitious goals of the German government in deployment of hydrogen, there are two main options[41] for infrastructure:

- injection of hydrogen into gas infrastructure
- building separate hydrogen pipelines or the conversion of parts of the gas infrastructure into hydrogen infrastructure.

[41] A third option, hydrogen chemically bonded with metal hydrides or liquid organic hydrogen carriers has not entered the market, yet.

Finally, possible paths for German gas imports and gas consumption are presented below. These examples demonstrate one of a few forecasting methods for *forecasts into the future*, without the possibility of computing accuracy measures. Although paths resemble scenarios in modelling analytical work of EIA or IEA, they were simulated with random value of error terms.

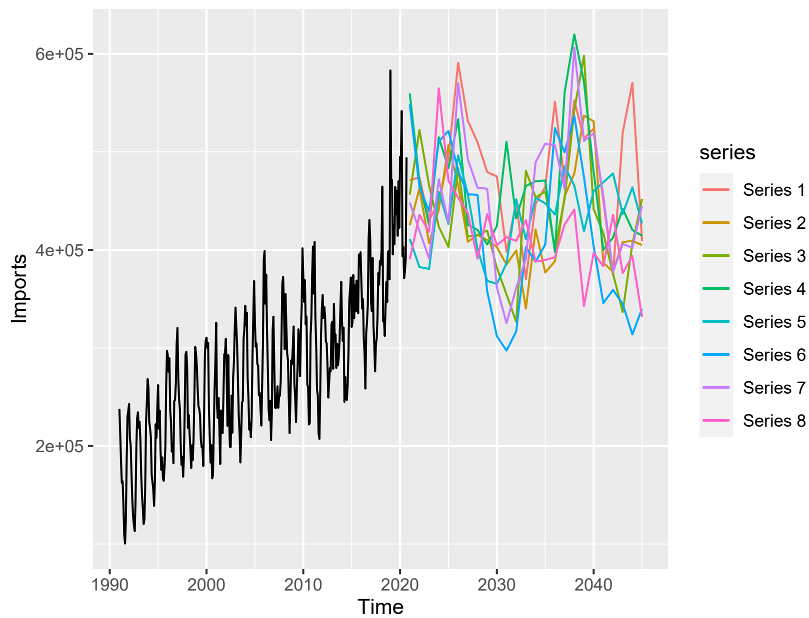


Figure 5.23: Paths for German gas imports for next 25 years. Nnetar function (non-linear autoregressive model) uses lagged values of the time series (Hyndman (2017)). Weights in "repeats" start with random values.

.

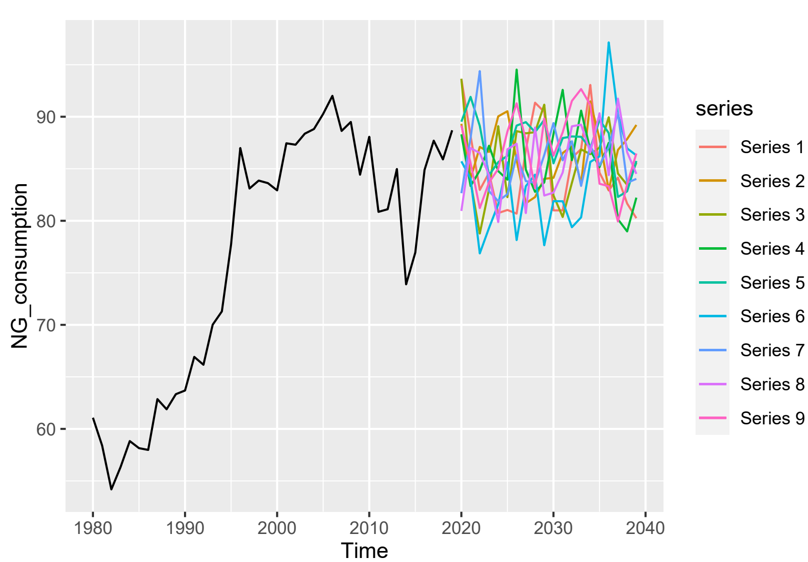


Figure 5.24: Nine paths for gas consumption in Germany. Although paths resemble scenarios in modelling analytical work of EIA or IEA, they were simulated with random value of error terms.

6 Conclusion

THE GOAL OF THIS WORK was to examine possibilities and trends in gas forecasting while understanding the principles behind commonly used models. First, forecasts based only on expert opinions prevailed.[1]

[1] The simple median or the mean of a few experts' forecasts would be in line with the outcome of the research paper of Mannes et al. (2014) titled "The Wisdom of Select Crowds". In current trends, forecast combinations make use of their work by considering single models as *expert opinions*.

In the IR field, after decades of efforts to use quantitative methods, expert analysis of events still dominates among methods.

Reviewing the research questions

- RQ1 *What are current approaches to predict gas imports and gas consumption at the country level? What is the impact of the domain knowledge of the gas sector on forecasting?*

  GAS IMPORTS have been rarely forecast, with reasons for this listed in chapter 4. Methods for forecasting gas consumption are listed and explained in chapter 3 with modelling exercises in chapters 4 and 5 for gas imports and consumption, respectively. These exercises provide insights into the current possibilities of forecasting packages[2] and are comparable with studies conducted by Beyca et al. (2019)[3] and Akpinar and Yumusak (2016).[4] Previous studies served as an inspiration in terms of searching for suitable input variables and measures of accuracy. The transferability of results of any study is low due to the unique setup of 1) data set (information sample of the forecasted process), 2) choice of models, 3) goal of a study, the length of a forecasting horizon and scope (company, city, state, global level) of forecasting.

  [2] Commercial forecasting tools have been mentioned but not directly used in these chapters.

  [3] Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., and Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, 80:937–949

  [4] Akpinar, M. and Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies*, 9(9):727

  DOMAIN KNOWLEDGE applies while simulating gas flow in nodes of a pipeline system and *prior to modelling* in the process of construction of a data set. Two data sets were necessary as short-term and long-term forecasting require specific input variables, the latter one including GDP, population, and price of alternative fuels, to name a few. For the long-term scenarios, some models in chapter 5 include the simulation of the newly gathered data to the model (ANN, one year data, previous year, true data).

THE FINAL SIZE OF BOTH DATA SETS depended on 1) fulfilling minimum requirements of a data set for a few models, 2) including all relevant variables if data was available from reliable sources, and 3) the selection of a starting point and the last known observations for time series without missing values. No forecasts have been adjusted as this would introduce *intentional* bias (Armstrong and Green (2017)[5].

[5] This work complies with most of the golden rules in the Forecasting checklist, introduced by Armstrong and Green (2017) but combination of forecasts.

WHY WERE NOT MORE EX ANTE FORECASTS PRODUCED? The forecast error would increase dramatically due to the uncertainty of the future values of inputs such as population and gas production in Germany (lower risk) or weather (higher uncertainty), and this would make prediction intervals too wide to make any statement about the future imports or consumption.[6]

Changes in population are small; the expected gas production in Germany up to 2030 is known in FNB Gas (2019), source: BVEG, e.V.

[6] Standard commercial software does not account for this higher uncertainty and keeps the width of the prediction interval the same over the periods (i.e. months, years).

- RQ2 *Is there any relation between the complexity of the data set, forecasting models and the process modelled (gas flows) to the accuracy of forecasts?*

SECTION 3.3 DISCUSSES various complexity criteria. In accordance with the literature on data complexity[7], regularity in data sets has been chosen as a measure for computing complexity of data sets (Sample Entropy, Approximate Entropy). The Approximate Entropy calculates the complexity of the process (model) whereas derived versions of Shannon entropy (information theory) calculate the complexity of a measure.

[7] "a pattern is simple if it can be generated by a short program or if it can be compressed, which essentially means that the pattern has some "regularity" in it." (Li and Abu-Mostafa (2006))

AS FOR THE COMPLEXITY OF FORECASTING MODELS, either the computational complexity (computing time necessary for a model to complete the task of forecasting), or definition based on understanding the model by uses are taken into account. For the latter definition, naïve or no-change model without seasonal adjustment, seasonal adjustment, single-exponential smoothing, Holt's exponential smoothing,dampened exponential smoothing and simple average of the exponential smoothing forecasts are regarded as simple (Green and Armstrong (2015)).

REGARDING COMPLEX PROCESSES TO BE MODELLED, in physics, any phenomenon is too complex and therefore physicists work with simplifications, set boundaries, and specified conditions until the phenomenon is computable. Present trends in forecasting have shown the opposite trend, as more complex models are introduced to probe new ways to forecast. Models of higher complexity go hand in hand with the complexity of modelled processes themselves (e.g. electricity production in hydro power plants).

FINALLY, deep inspection of the relation of complexity towards the data set[8], models and processes modelled would require avoiding the connection to energy and forecasting as these relations are supposed to be valid in general terms.

[8] For instance, searching for the input-output relationship present in the training data set as noted in Li and Abu-Mostafa (2006)).

6.1 *Closing thoughts*

- More accurate forecasts provide marginal contributions for the gas sector or for energy policy, if they are neglected while making decisions at the gas company or state level. Ministries give assignments to institutes to model situations as if climate goals were achieved by 2050. These aspirational forecasts have to be distinguished from forecasting based on data. Second, policy makers, as most people, prefer a narrower interval, even without the true value included, than the wider interval, which they perceive as too wide (Yaniv and Foster (1995)). In comparison to finely grained judgements, imprecise judgements are less informative and less attention is paid to them from experts outside the forecasting community; however, they are likely to be more accurate.
- Contextual information on commodities such as *oil, gas or solar technology is* strictly bound to the existing available infrastructure and its capacity factors. In turn, this enables the final reality-check of forecasts.
- Forecasting, especially with ML methods, is more vital for higher frequencies; higher than hourly values approaching even real-time data forecasting. By this, gas demand forecasting at the state level cannot reap the advantages of new trends as data of higher frequencies is available only to in-house experts in charge of their market area. Data sets used in chapters 4 and 5 contain low resolution data (monthly, yearly) which can decrease accuracy of results (Hong et al. (2014)). The sample (every data set represents just a sample of information on the process) is insufficient for building an appropriate model, therefore features of sophisticated models cannot be utilized to their full extent. Therefore, explanatory variables as well as contextual information was included. In their latest paper, Makridakis et al. (2020)[9] conclude explanatory variables can be useful if there are accurate forecasts of the explanatory variables

[9] Makridakis, S., Hyndman, R. J., and Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, 36(1):15–28. https://doi.org/10.1016/j.ijforecast.2019.05.011

available (e.g. in the case of German gas production) and when assumed relationships between an output and explanatory variables probably continue into the future. The latter condition is met for weather forecasts. In electricity load forecasting, the strong correlation between the temperature and load is widely accepted in the field. However, this is valid only for the US or other countries in warm climates with widespread air-conditioning use, meaning an extensive use of electric heating and/or cooling units. Therefore, *in energy forecasting*, data sets and assumed relationships between a forecast and explanatory variables must be country-specific.

THE TREMENDOUS NUMBER OF STUDIES on (electricity) load forecasting, hydro power forecasting, and political forecasting shows an imperfect transferability of methods[10] and the wish to test current models from new perspectives. Amendments take a form of stacking and shuffling data sets from various countries for forecasting electricity prices, direct forecasting of peak load from data on temperature, etc. Moreover, full-time forecasters prefer producing synthetic data sets to control the quality of probabilistic forecasts (bias, error, variance, covariance) and enhance their understanding of a models' performances.

[10] In classical forecasting problems, researchers compare models as presented in chapters 4 and 5, based on the accuracy measure(s) of their choice, and identify the model that outperforms other models. However, no model can show outstanding performance for all data sets and forecasting problems.

6.2 *The value of forecasting*

THIS WORK HAS NOT TOUCHED environmental aspects of gas use, especially flaring, methane emissions during the distribution, boil-off during the LNG transport and emissions from gas combustion. Although constraints on emissions are built-in in constructed energy scenarios, they matter in developed countries only. Rather, this research drew attention to *assumptions behind forecasting and modelling*, which should serve a better understanding of their products (i.e. forecasts, scenarios) and, thus, using them with caution in energy planning.

BUSINESS FORECASTING MEASURES VALUE in monetary units and poses the question of "what is the actual economic value of forecasts?" Forecasting in the environmental sciences does not provide straightforward benefits; it does not save the environment nor can it prevent harmful acts due to inefficient energy supply planning. However, an improved ability to accurately anticipate energy consumption, emissions, or inflation rates, would benefit decision-making. For this to happen, ever improving methodologies for the anticipation must be applied.

FORECASTS ASSIST INVESTMENT, trading decisions and policies on energy security. Although all forecasts are incorrect, it is better to plan actions now based on incorrect forecasts than to plan without forecasts at all. Post 2030, it will not be macro-economic variables (e.g. GDP or, population) that will determine future gas demand in Germany, but rather policies of phasing-out natural methane-based gases in Germany as well as in neighbouring countries.

Alphabetical Index

List of Symbols

n number of data points. 35, 37, 48

p the number of estimated parameters in the model. 47, 51, 66

r_t relative error. 35

r Pearson's correlation coefficient. 36

τ training data set. 34

x predictor variable, an observation in the data set. 45, 100

y_t actual value. 35

y prediction, outcome, output, target variable. 34, 45, 48, 65, 100

Appendices

Input	Unit	Description	Source
Heating Degree Days	-	Germany	Eurostat
Cooling Degree Days	-	Germany	Eurostat
Imported Gas	TJ		BAFA Website:`https://www.bafa.de`
Cross-border Price	EUR/TJ		BAFA `https://www.bafa.de`
Gas consumption for Electricity Production	GJ	Germany	Destatis.de, Genesis-Tabelle 43311-0002, `https://www-genesis.destatis.de`
Domestic Natural Gas Production	TJ		W. E. G. Wirtschaftsverband Erdöl- und Erdgasgewinnung e. V.
Gas - Export	TJ	Considered but Excluded	BAFA
Gas Storage Balance	TJ		BAFA

Table 1: Data key for German gas imports

Table 2: Data key for German gas consumption

Variable Name	Meaning	Unit	Source and Notes
Year	Year of data collection; denotes a single observation	Numerical	
GDP	Gross domestic product (Germany), constant prices	Billions of National Currency (Euros)	IMF, World Economic Outlook Database (WEO), original source: National Statistics Office. Notes: Data until 1990 refers to German federation only (West Germany). Base year: 2010. The GDP deflator in the base year is not exactly equal to 100 since it is computed based on the quarterly real GDP data, which have been adjusted for seasonal and calendar effects. Type: Expenditure-based GDP.
Pop	Population of Germany	Millions of Persons	IMF, WEO, April 2019, original source: National Statistics Office. Latest actual data: 2018 Notes: Data until 1990 refers to German federation only (West Germany). Data from 1991 refer to United Germany. Data last updated: 03/2019
NG_Res	Proved Natural gas reserves (Germany)	Trillion Cubic Meters	BP from the Data product: Energy Production and Consumption. Permalink: `https://www.quandl.com/data/BP/GAS_RESERVES_DEU` Quandl Code: BP/GAS_RESERVES_DEU Description: Proved reserves of natural gas - Generally taken to be those quantities that geological and engineering information indicates with reasonable certainty can be recovered in the future from known reservoirs under existing economic and operating conditions.For the 2018, data are taken from: LBEG, Landesamt für Bergbau, Energie und Geologie. Oil and gas reserves in Germany on 1st January 2019 (In German: Erdöl- und Erdgasreserven in der Bundesrepublik Deutschland am 1. Januar 2019, Tab. 5: Rohgasreserven am 1.1.2019 nach Fördergebieten (in Mrd. m^3(Vn)), Link: www.niedersachsen.de Last check: 3rd July 2020. Observations from the first source and the second one for 2018 is equal.

Continued on next page

Table 2 – continued from previous page

Variable Name	Meaning	Unit	Source and Notes
NG_Prod	NG production (Germany)	Billion Cubic Meters	BP from the Data product: Energy Production and Consumption. permalink: https://www.quandl.com/data/BP/GAS_PROD_DEU Quandl Code: BP/GAS_PROD_DEU.As far as possible, the data represents standard cubic metres measured at 15C and 1013 millibar (mbar); as they are derived directly from tonnes of oil equivalent using an average conversion factor, they do not necessarily equate with gas volumes expressed in specific national terms.
Brent_Price	Average annual Brent crude oil	U.S.dollars/barrel	OPEC; IEA. Published by: MWV.de, ID 262860, June 2019. Link: https://www.statista.com/statistics/262860/uk-brent-crude-oil-price-changes-since-1976/ Original source: mwv.de, ID 262860
DE_Oil_Price	Price of oil for Germany (Western, change to whole in 1999)	Euros per Hectoliter	www.destatis.de. Price for light heating oil. In German: Preise. Erzeugerpreise gewerblicher Produkte (Inlandabsatz). Preise für leichtes Heizöl, Motorenbenzin und Dieselkraftstoff. Lange Reihen ab 1976 bis Mai 2019, Artikelnummer: 561240219054
Elect_Gen	Electricity generation (Germany)	TWh	BP from the Data product: Energy Production and Consumption. Permalink: https://www.quandl.com/data/BP/ELEC_GEN_DEU Quandl code: BP/ELEC_GEN_DE Based on the gross output.Value for 2018: https://www.umweltbundesamt.de/sites/_default/files/medien/384/bilder/dateien/2_datentabelle-zur-abb_entw-bruttostromerzeugung-verbrauch_2019-02-26.pdf
NG_Imp	Natural gas imports	TJ	International Energy Agency, from 1999 on: Bundesamt für Wirtschaft und Ausfuhrkontrolle (BAFA)
Heat_DE	Number of heating degree days in Germany	Numerical	EUROSTAT. Accessed on 30th June 2019. Until 1990 – BRD.

Continued on next page

Table 2 – continued from previous page

Variable Name	Meaning	Unit	Source and Notes
Cool_DE	Number of cooling degree days in Germany	Numerical	EUROSTAT. Accessed on 30th June 2019. Until 1990-BRD.
Avg_Price_In	Average price of NG (industry) (Germany)	Cent/kWh	www.destatis.de. Excluded: Value-added tax (VAT) Included: usage charges (in German: Nutzungsentgelte), gas taxes
Avg_Price_H	Average price of NG (household/domestic) (Germany)	Cent/kWh	www.destatis.de. Excluded: Value-added tax (VAT). Included: usage charges (in German: Nutzungsentgelte), gas taxes
Avg_Price	Total average price of NG (Germany)	Cent/kWh	www.destatis.de. Excluded: Value-added tax (VAT). Included: usage charges (in German: Nutzungsentgelte), gas taxes
Oil_Con	Oil Consumption	Consumption in million tonnes oil equivalent*	BP, Statistical Review of World Energy, 2019, 68th Edition.https://www.bp.com/content/-dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf
Hydro_E	Hydro electricity	TWh	BP, Statistical Review of World Energy https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html
Biomass	Electricity produced by biomass	TWh	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html

Continued on next page

Table 2 – continued from previous page

Variable Name	Meaning	Unit	Source and Notes
Solar	Electricity produced by solar energy	TWh	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html
Wind	Electricity produced from wind energy	TWh	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html
Coal_Con	Coal consumption	MTOE	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html
Nuclear	Electricity from nuclear power plants	TWh	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html
NG_Con	Natural gas consumption (Germany)	Billion Cubic Meters	BP. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/downloads.html

7 Bibliography

Akpinar, M. and Yumusak, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series methods. *Energies*, 9(9):727.

Al-Fattah, S. and Startzman, R. (1999). Analysis of world natural gas production. Conference: SPE Eastern Regional Meeting, paper SPE-57463. Charleston, West Virginia, USA. https://doi.org/10.2118/57463-MS Last accessed 2021-03-26.

Altfeld, K. and Pinchbeck, D. (2013). Admissible hydrogen concentrations in natural gas systems. *Gas for Energy*, (03). https://www.gerg.eu/wp-content/uploads/2019/10/HIPS_Final-Report.pdf Last accessed 2020-12-19.

Anagnostis, A., Papageorgiou, E., Dafopoulos, V., and Bochtis, D. D. (2019). Applying long short-term memory networks for natural gas demand prediction. In Bourbakis, N. G., Tsihrintzis, G. A., and Virvou, M., editors, *10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019, Patras, Greece, July 15-17, 2019*, pages 1–7. IEEE. https://doi.org/10.1109/IISA.2019.8900746 Last accessed 2020-08-31.

Anastasiadis, A. D., Magoulas, G. D., and Vrahatis, M. N. (2005). New globally convergent training scheme based on the resilient propagation algorithm. *Neurocomputing*, 64:253–270. http://www.sciencedirect.com/science/article/pii/S0925231204005168 Last accessed 2020-09-06.

Andrade-Cabrera, C., de Rosa, M., Kathirgamanathan, A., Kapetanakis, D.-S., and Finn, D. (2018). A study on the trade-off between energy forecasting accuracy and computational complexity in lumped parameter building energy models. The 10th Canada Conference of International Building Performance Simulation Association (eSim 2018), Montreal, Canada, 9-10 May 2018. http://hdl.handle.net/10197/9663 Last accessed 2019-03-22.

Ang, B. W., Choong, W. L., and Ng, T. S. (2015). Energy security: Definitions, dimensions and indexes. *Renewable and Sustainable Energy Reviews*, 42:1077–1093.

Armstrong, J. and Green, K. (2017). Forecasting methods and principles: Evidence-based checklists. Working paper. https://doi.org/10.1080/21639159.2018.1441735 Last accessed 2021-03-28.

Ascher, W. (1978). *Forecasting: An appraisal for policy makers and planners*. Johns Hopkins University Pr, Baltimore. ISBN 0801820359.

Asdal, K. (2011). The office: The weakness of numbers and the production of non-authority. *Accounting, Organizations and Society*, 36(1):1–9.

Astakhova, O. (2019). 12.10. 2019 aramco hopes to repair remaining damage from attacks by end-November. reuters. https://www.reuters.com/article/us-saudi-aramco-attacks/aramco-hopes-to-repair-remaining-damage-from-attacks-by-\end-november-idUSKBN1WR0JS Last accessed 2021-04-08.

Atiya, A. F. (2020). Why does forecast combination work so well? *International Journal of Forecasting*, 36(1):197–200.

Bale, C. S., Varga, L., and Foxon, T. J. (2015). Energy and Complexity: New ways forward. *Applied Energy*, 138:150–159. https://doi.org/10.1016/j.apenergy.2014.10.057.

Barker, J. (2020). Machine learning in m4: What makes a good unstructured model? *International Journal of Forecasting*, 36(1):150–155.

Baughman, K. A., Cecchi, G. A., Kozloski, R. J., and Brian, O. M. (1988). Selecting Forecasting Model Complexity using eigenvalues. us patent 10,469,398. https://patents.google.com/patent/US10469398B2/en International Business Machines Corporation.Last accessed 2021-04-01.

Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., and Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics*, 80:937–949.

Bezdek, R. and Wendling, R. (2002). A half century of long-range energy forecasts: Errors made, lessons learned, and implications for forecasting. *Journal of Fusion Energy*, 21. https://doi.org/10.1023/A:1026208113925.

Biresselioglu, M. E., Yelkenci, T., and Oz, I. O. (2015). Investigating the natural gas supply security: A new perspective. *Energy*, 80:168–176.

BMWi (2019). Monitoring-Bericht nach § 51 EnWG zur Versorgungssicherheit bei Erdgas. Stand: Februar 2019. Berlin. https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/monitoringbericht-versorgungssicherheit-2017.html Last accessed 2020-12-19.

Botev, L. and Johnson, P. (2020). Applications of statistical process control in the management of unaccounted for gas. *Journal of Natural Gas Science and Engineering*, 76:103194.

BP (2019). BP Statistical Review of World Energy. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf Last accessed 2020-12-19.

BP (2020). BP Statistical Review of World Energy. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-natural-gas.pdf Last accessed 2020-12-21.

Bradshaw, M. and Boersma, T. (2020). *Natural Gas*. Resources. Wiley. ISBN 9780745659978, LCCN 2019050902.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32. https://doi.org/10.1023/A:1010933404324.

Bridge, G. and Bradshaw, M. (2017). Making a global gas market: Territoriality and production networks in liquefied natural gas. *Economic Geography*, 93(3):215–240.

Browell, J. (2015). *Spatio-temporal prediction of wind fields*. PhD thesis, University of Strathclyde. http://oleg.lib.strath.ac.uk:80/R/?func=dbin-jump-full&object_id=25822 Last accessed 2020-12-20.

Bundesnetzagentur and Bundeskartellamt (2019). Monitoringbericht gemäß § 63 Abs. 3 i. v. m. § 35 EnWG und § 48 abs. 3 i. v. m. § 53 Abs. 3 GWB Stand: 29. Mai 2019. https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Monitoringberichte/Monitoringbericht2018.pdf;jsessionid=7947A2EB1993EAEF55BD446D286E487B?__blob=publicationFile&v=6 Last accessed 2019-10-20.

Busse, S., Helmholz, P., and Weinmann, M. (2012). Forecasting day ahead spot price movements of natural gas - an analysis of potential influence factors on basis of a NARX neural network. *Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband der MKWI 2012*. https://publikationsserver.tu-braunschweig.de/servlets/

MCRFileNodeServlet/dbbs_derivate_00027726/Beitrag299.pdf Last accessed 2021-04-14.

Cage, J. (1952). 4'33 Piano solo arrangement. First movement. https://musescore.com/user/5832946/scores/1559096 Last accessed 2020-07-26.

Chatfield, C. (2001). *Time-series forecasting*. Chapman & Hall/CRC, Boca Raton. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10143032 Last accessed 2020-12-20.

Chen, J. (2002). An entropy theory of value. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.307442.

Chen, J. (2018). An entropy theory of value. *Structural Change and Economic Dynamics*, 47:73–81.

Chen, Y., Chua, W. S., and Koch, T. (2018). Forecasting day-ahead high-resolution natural gas demand and supply in Germany. *Applied Energy*, 228:1091–1110.

Chkili, W., Hammoudeh, S., and Nguyen, D. K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics*, 41:1–18.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Darin, S. G. and Stellwagen, E. (2020). Forecasting the M4 competition weekly data: Forecast pro's winning approach. *International Journal of Forecasting*, 36(1):135–141.

de Nadai, M. and van Someren, M. (2015). Short--term anomaly detection in gas consumption through ARIMA and Artificial Neural Network forecast. In *2015 IEEE Workshop on Environmental, Energy and Structural Monitoring Systems (EESMS 2015)*, pages 250–255. IEEE. ISBN 978-1-4799-8215-8.

Debnath, K. B. and Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, 88:297–325. https://doi.org/10.1016/j.rser.2018.02.002.

Dey, H. S., Kabir, M. A., Wadud, Z., Khan, S. I., and Azad, M. A. K. (2011). Econometric modeling and forecasting of natural gas demand for power sector in Bangladesh. IEEE Region 10 Annual International Conference, Proceedings/TENCONISBN 978-1-4577-0256-3.

Dimitriadou, A., Gogas, P., Papadimitriou, T., and Plakandaras, V. (2018). Oil market efficiency under a machine learning perspective. *Forecasting*, 1(1):157–168.

DNV-GL (2019). Energy Transition Outlook. Oil and Gas Forecast 2050. https://eto.dnv.com/2017/oilgas Last accessed 2021-03-26.

Dreborg, H. K. (2004). *Scenarios and Structural Uncertainty: Explorations in the Field of Sustainable Transport*. Trita-INFRA. ISBN 91-7323-068-5.

E. E. Holmes, M. D. Scheuerell and E. J. Ward (2020). 4.4 Correlation within and among time series | Applied Time Series Analysis for Fisheries and Environmental Sciences. https://nwfsc-timeseries.github.io/atsa-labs/sec-tslab-correlation-within-and-among-time-series.html Last accessed 2020-12-20.

ENTSOG (2017). Security of Supply Simulation Report. https://www.entsog.eu/security-of-supply-simulation Last accessed 2021-04-14.

ENTSOG (2019). The European Natural Gas Network: Capacities at cross-border points on the primary market. https://www.entsog.eu/sites/default/files/2020-01/ENTSOG_CAP_2019_A0_1189x841_FULL_401.pdf Last accessed 2020-10-29.

Erdoğdu, E. (2010). Natural gas demand in Turkey. *Applied Energy*, 87(1):211–219. https://doi.org/10.1016/j.apenergy.2009.07.006.

Ervural, B. C., Beyca, O. F., and Zaim, S. (2016). Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption. *Procedia - Social and Behavioral Sciences*, 235(235):537–545. https://doi.org/10.1016/j.sbspro.2016.11.066.

Eser, P., Chokani, N., and Abhari, R. (2019). Impact of Nord Stream 2 and LNG on gas trade and security of supply in the European gas network of 2030. *Applied Energy*, 238:816–830. https://doi.org/10.1016/j.apenergy.2019.01.068.

EU (2012). Eur-lex-12012e: Consolidated version of the Treaty on the Functioning of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012E/TXT Last accessed 2019-07-22.

Fallah, S., Ganjkhani, M., Shamshirband, S., and Chau, K.-w. (2019). Computational intelligence on short-term load forecasting: A methodological overview. *Energies*, 12(3):393.

Federal Ministry for Economic Affairs and Energy (2019). Preventive Action Pla for Gas for the Federal Republic of Germany. https://ec.europa.eu/energy/sites/ener/files/documents/2019.06.21-de-pap-praventionsplan_englisch_final_1.pdf Last accessed 2020-12-27.

Fischer, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368.

FNB Gas (June 2019). Netzentwicklungsplan Gas 2020-2030 (Grid Development Plan): Szenariorahmen. https://www.fnb-gas.de/media/fnb_gas_2020_1_sr_konsultation_de_kf_1.pdf Last accessed 2020-12-20.

Fortes, P., Alvarenga, A., Seixas, J., and Rodrigues, S. (2015). Long-term energy scenarios: Bridging the gap between socio-economic storylines and energy modeling. *Technological Forecasting and Social Change*, 91:161–178.

Fragkos, P., Kouvaritakis, N., and Capros, P. (2015). Incorporating uncertainty into world energy modelling: the PROMETHEUS model. *Environmental Modeling & Assessment*, 20(5):549–569.

Franco, A. and Fantozzi, F. (2015). Analysis and clustering of natural gas consumption data for thermal energy use forecasting. *Journal of Physics: Conference Series*, 655:012020.

Friedrichs, J. (2013). *The future is not what it used to be: Climate Change and Energy Scarcity*. MIT Press, Cambridge, Mass. ISBN 9780262019248.

Gaba, A., Tsetlin, I., and Winkler, R. L. (2017). Combining interval forecasts. *Decision Analysis*, 14(1):1–20.

Gas Infrastructure Europe (2021). Lng Import Terminals Map Database May 2019. https://www.gie.eu/index.php/gie-publications/databases/lng-database Last accessed 2021-04-13.

Ghalehkhondabi, I., Ardjmand, E., Weckman, G. R., and Young, W. A. (2017). An overview of energy demand forecasting methods published in 2005–2015. *Energy Systems*, 8:411–447. 10.1007/s12667-016-0203-y.

Ghandar, A., Michalewicz, Z., and Zurbruegg, R. (2016). The relationship between model complexity and forecasting performance for computer intelligence optimization in finance. *International Journal of Forecasting*, 32(3):598–613.

Gianfreda, A., Ravazzolo, F., and Rossini, L. (2020). Comparing the forecasting performances of linear models for electricity prices with high RES penetration. *International Journal of Forecasting*, 36(3):974–986.

Gillessen, B., Heinrichs, H., Hake, J.-F., and Allelein, H.-J. (2019). Natural gas as a bridge to sustainability: Infrastructure expansion regarding energy security and system transition. *Applied Energy*, 251:113377.

González-Espinoza, A., Martinez-Mekler, G., and Lacasa, L. (2020). Arrow of time across five centuries of classical music. *Physical Review Research*, 2:33166.

Goodwin, P. (2020). Could these recent findings improve your judgmental forecasts? *Foresight: The International Journal of Applied Forecasting*, (56):7–9.

Grajcar, M. (2019). Energy Forecasting vs Energy Modelling. jazz improvisation vs Symphony. http://cyseni.com/archives/proceedings/Proceedings_of_CYSENI_2019.pdf Last accessed 2020-12-20.

Green, K. C. and Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*, 68(8):1678–1685.

Grüger, F., Hoch, O., Hartmann, J., Robinius, M., and Stolten, D. (2019). Optimized electrolyzer operation: Employing forecasts of wind energy availability, hydrogen demand, and electricity prices. *International Journal of Hydrogen Energy*, 44(9):4387–4397.

Günther, F. and Fritsch, S. (2010). Neuralnet: Training of neural networks. *R Journal*, 2(1):30–39. https://journal.r-project.org/archive/2010/RJ-2010-006/RJ-2010-006.pdf Last accessed 2020-12-20.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data mining, inference, and prediction*. Springer series in statistics. Springer, New York, second edition. ISBN 978-0-387-84857-0.

Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J.-J., and Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy*, 179:214–221.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10.

Hong, T. (06/30/2014). Global energy forecasting competition. past, present and future. https://forecasters.org/wp-content/uploads/gravity_forms/7-2a51b93047891f1ec3608bdbd77ca58d/2014/07/HONG_TAO_ISF2014.pdf Last accessed 2020-12-20.

Hong, T., Wilson, J., and Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, 5(1):456–462. https://doi.org/10.1109/TSG.2013.2274373.

Hong, W. S. (2013). *Intelligent energy demand forecasting*, volume 10 of *Lecture notes in energy*. Springer, London and Heidelberg and New York and Dordrecht. ISBN 9781447149682.

Hribar, R., Potočnik, P., Šilc, J., and Papa, G. (2019). A comparison of models for forecasting the residential natural gas demand of an urban area. *Energy*, 167:511–522.

Hyndman, R. (2017). nnetar. Neural Network Time Series Forecasts. https://www.rdocumentation.org/packages/\protect\discretionary{\char\hyphenchar\font}{}{}forecast/versions/8.13/topics/nnetar Last accessed 2021-01-10.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(1):1–22.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Hyndman, R. J. and Kostenko Andrey V. (2007). Minimum sample size requirements for seasonal forecasting models. *Foresight*, (6):12–15. https://robjhyndman.com/papers/shortseasonal.pdf.

International Energy Agency (2020). Germany 2020. Energy Policy Review. https://www.bmwi.de/Redaktion/DE/Downloads/G/germany-2020-energy-policy-review.pdf?__blob=publicationFile&v=4 Last accessed 2020-12-27.

International Energy Agency (2021). Energy security. https://www.iea.org/topics/energy-security Last accessed 2021-01-09.

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., and Callot, L. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, 36(1):167–177.

Jefferson, M. (2016). Energy realities or modelling: Which is more useful in a world of internal contradictions? *Energy Research & Social Science*, 22:1–6. https://doi.org/10.1016/j.erss.2016.08.006.

Jensterle, M., Narita, J., Piria, R., Samadi, S., Prantner, M., Crone, K., Siegemund, S., Kan, S., Matsumoto, T., Shibata, Y., and Thesen, J. (2019). The role of clean hydrogen in the future energy systems of Japan and Germany. Berlin: adelphi. https://www.adelphi.de Last accessed 2020-12-28.

Kaposty, F., Kriebel, J., and Löderbusch, M. (2020). Predicting loss given default in leasing: A closer look at models and variable selection. *International Journal of Forecasting*, 36(2):248–266. https://doi.org/10.1016/j.ijforecast.2019.05.009.

Karabiber, O. A. and Xydis, G. (2020). Forecasting day-ahead natural gas demand in Denmark. *Journal of Natural Gas Science and Engineering*, 76:103193. https://doi.org/10.1016/j.jngse.2020.103193.

Katz, J. H. (2020). Monitoring Forecast Models Using Control Charts. *Foresight: The International Journal of Applied Forecasting*, (56):20–25.

Keane, M. and Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *American Economic Review*, 80(4):714–35. https://EconPapers.repec.org/RePEc:aea:aecrev:v:80:y:1990:i:4:p:714-35.

Keuzenkamp, H. A. and McAleer, M. (1997). The complexity of simplicity. *Mathematics and Computers in Simulation*, 43(3-6):553–561.

Kuhn, M. and Johnson, K. (2016). *Applied predictive modeling*. Springer, New York, Corrected 5th printing edition. ISBN 978-1-4614-6849-3.

Lewis, C. (1982). *Industrial and business forecasting methods. A practical guide to exponential smoothing and curve fitting*. Butterworth, London. ISBN 0408005599.

Li, L. and Abu-Mostafa, Y. (2006). Data complexity in machine learning. https://resolver.caltech.edu/CaltechCSTR:2006.004 Computer Science Technical Reports. California Institute Of Technology, Pasedena, USA. Unpublished.

Li, Z. and Zhang, Y.-K. (2008). Multi-scale entropy analysis of Mississippi river flow. *Stochastic Environmental Research and Risk Assessment*, 22(4):507–512. https://doi.org/10.1007/s00477-007-0161-y.

Lichtendahl, K. C. and Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, 36(1):142–149. https://doi.org/10.1016/j.ijforecast.2019.03.027.

Lindley, D. V. (2001). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):293–337. https://doi.org/10.1111/1467-9884.00238.

Liu, S., editor (2011). *Proceedings of 2011 IEEE International Conference on Grey Systems and Intelligent Services (GSIS) with the 15th WOSC International Congress on Cybernetics and Systems, Nanjing, China, 15 - 18 September 2011*, Piscataway, NJ. IEEE.

Livni, R., Shalev-Shwartz, S., and Shamir, O. (2014). On the computational efficiency of training neural networks. https://arxiv.org/pdf/1410.1141 Last accessed 2020-12-20.

Lund, H., Arler, F., Østergaard, P., Hvelplund, F., Connolly, D., Mathiesen, B., and Karnøe, P. (2017). Simulation versus optimisation: Theoretical positions in energy system modelling. *Energies*, 10(7):840. https://doi.org/10.3390/en10070840.

Ma, Y. and Li, Y. (2010). Analysis of the supply-demand status of China's natural gas to 2020. *Petroleum Science*, 7(1):132–135. https://link.springer.com/content/pdf/10.1007/s12182-010-0017-9.pdf.

Magdowski, A. and Kaltschmitt, M. (2017). Dependency of forecasting accuracy for balancing power supply by weather-dependent renewable energy sources. *Jordan Journal of Mechanical and Industrial Engineering*, 11:209–215.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4):527–529. https://doi.org/10.1016/0169-2070(93)90079-3.

Makridakis, S. (1995). Forecasting accuracy and system complexity. *RAIRO - Operations Research - Recherche Opérationnelle*, 29(3):259–283. http://www.numdam.org/item/RO_1995__29_3_259_0/.

Makridakis, S., Hyndman, R. J., and Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *International Journal of Forecasting*, 36(1):15–28. https://doi.org/10.1016/j.ijforecast.2019.05.011.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808.

Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and applications*. Wiley, Hoboken, NJ, 3. ed. edition. ISBN 9780471532330.

Mannes, A., Soll, J., and Larrick, R. (2014). The wisdom of select crowds. *Journal of personality and social psychology*, 107:276–299.

May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467.

Melikoglu, M. (2013). Vision 2023: Forecasting Turkey's natural gas demand between 2013 and 2030. *Renewable and Sustainable Energy Reviews*, 22:393–400.

Merkel, G., Povinelli, R., and Brown, R. (2018). Short term load forecasting of natural gas with deep neural network regression. *Energies*, 11(8).

Merkel G. D. et al. (2017). Deep neural network regression for short-term load forecasting of natural gas. *Proceedings of the International Symposium on Forecasting*. https://epublications.marquette.edu/electric_fac/287 Electrical and Computer Engineering Faculty Research and Publications. 287.

Mevik, B.-H. and Cederkvist, H. R. (2004). Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *Journal of Chemometrics*, 18(9):422–429.

Miller, A. (2006). Gazprom. Strategy for the energy sector leadership. Speech by Alexey Miller at the Annual General Shareholders Meeting . https://www.gazprom.com/press/news/miller-journal/2006/99967/ Last accessed 2021-03-25.

Ministry for Economic Affairs, F. and Energy (2020). The National Hydrogen Strategy. https://www.bmbf.de/files/bmwi_Nationale$%$20Wasserstoffstrategie_Eng_s01.pdf Last accessed 2020-12-28.

Müller-Syring, G., Henel, M., Henel, M., Köppel, W., Mlaker, H., Sterner, M., and Höcher, T. (2013). DVGW. Entwicklung von modularen Konzepten zur Erzeugung, Speicherung und Einspeisung von Wasserstoff und Methan ins Erdgasnetz. https://www.dvgw.de/medien/dvgw/forschung/berichte/g1_07_10.pdf Last accessed 2019-07-25.

Nuttall, W. J. and Manz, D. L. (2008). A new energy security paradigm for the twenty-first century. *Technological Forecasting and Social Change*, 75(8):1247–1259.

Nymoen Strategieberatung (2016). Strategische Marktprognose Erdgas. https://zukunft.erdgas.info/fileadmin/public/PDF/Politischer_Rahmen/marktprognose-erdgas-2016.pdf Last accessed 2019-12-27.

Overholt, W. H. (2000). Forecasting: An Appraisal for Policymakers and Planners. *Policy Sciences*, 33(1):101–106.

Pawlik V. (2019). Gasanschluss im Haushalt in Deutschland 2018 | Statista. https://de.statista.com/statistik/daten/studie/181631/umfrage/gasanschluss-im-haushalt-vorhanden/ Last accessed 2019-11-09.

Pelikán, E. (2014). Forecasting of processes in complex systems for real-world problems. Tutorial. *Neural Network World*, 24(6):567–589. http://www.nnw.cz/doi/2014/NNW.2014.24.032.pdf.

Petropoulos, F., Hyndman, R. J., and Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554.

Pincus, S. (2008). Approximate entropy as an irregularity measure for financial data. *Econometric Reviews*, 27(4-6):329–362. https://doi.org/10.1080/07474930801959750.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88(6):2297–2301.

Potočnik, P., Soldo, B., Šimunović, G., Šarić, T., Jeromen, A., and Govekar, E. (2014). Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia. *Applied Energy*, 129:94–103.

Prehn, O. (11/08/2018). Ten steps to improvise jazz. https://www.youtube.com/watch?v=BgvAq1tTnN4&t=1778s Last accessed 2020-05-12.

Pulhan, A., Yorucu, V., and Sinan Evcan, N. (2020). Global energy market dynamics and natural gas development in the Eastern Mediterranean region. *Utilities Policy*, 64:101040.

R Documentation (2020). Holt-Winters function | R documentation. https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/HoltWinters Last accessed 2020-09-06.

Ramesh, K. S. (2011). The deterministic chaos in heart rate variability signal and analysis techniques. *International Journal of Computer Applications*, 35:39–46.

Raza, M. Q. and Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352–1372.

Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology. Heart and circulatory physiology*, 278(6):H2039–49.

Rieck, B. A. (2017). *Persistent Homology in Multivariate Data Visualization*. PhD thesis, Heidelberg University Library. http://archiv.ub.uni-heidelberg.de/volltextserver/22914/1/Dissertation.pdf Last accessed 2021-04-13.

Rácz, L. and Németh, B. (2018). Investigation of dynamic electricity line rating based on neural networks. *Energetika*, 64(2).

Scarpa, F. and Bianco, V. (2017). Assessing the quality of natural gas consumption forecasting: An application to the Italian residential sector. *Energies*, 10(11):1879.

Scher, S. and Messori, G. (2019). Weather and climate forecasting with neural networks : using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809. https://doi.org/10.5194/gmd-12-2797-2019.

Semenychev, V. K., Kurkin, E. I., and Semenychev, E. V. (2014). Modelling and forecasting the trends of life cycle curves in the production of non-renewable resources. *Energy*, 75:244–251.

Sen, D., Günay, M. E., and Tunç, K. M. (2019). Forecasting annual natural gas consumption using socio-economic indicators for making future policies. *Energy*, 173:1106–1118.

Sethna, P. James (2020). *Entropy, Order Parameters, and Complexity*. Clarendon Press, Oxford. http://pages.physics.cornell.edu/~sethna/StatMech/EntropyOrderParametersComplexity20.pdf Last accessed 2021-04-13.

Shaikh, F. and Ji, Q. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *International Journal of Electrical Power & Energy Systems*, 77:25–32.

Sharmina, M., Abi Ghanem, D., Browne, A. L., Hall, S. M., Mylan, J., Petrova, S., and Wood, R. (2019). Envisioning surprises: How social sciences could help models represent 'deep uncertainty' in future energy and water demand. *Energy Research & Social Science*, 50:18–28. https://doi.org/10.1016/j.erss.2018.11.008.

Siemek, J., Nagy, S., and Rychlicki, S. (2003). Estimation of natural gas consumption in Poland based on the logistic-curve interpretation. *Applied Energy*, 75(1-2):1–7.

Smil, V. (2000). Perils of long-range energy forecasting. *Technological Forecasting and Social Change*, 65(3):251–264.

Soldo, B. (2012). Forecasting natural gas consumption. *Applied Energy*, 92:26–37. https://doi.org/10.1016/j.apenergy.2011.11.003.

Sovacool, B. K., Ryan, S. E., Stern, P. C., Janda, K., Rochlin, G., Spreng, D., Pasqualetti, M. J., Wilhite, H., and Lutzenhiser, L. (2015). Integrating social science in energy research. *Energy Research & Social Science*, 6:95–99.

Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. *Energy*, 178:585–597.

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85:208–220.

Tamba, J. G., Essiane, S. N., Sapnken, E. F., Koffi, F. D., Nsouandélé, J. L., Soldo, B., and Njomo Donatien (2018). Forecasting natural gas: A literature survey. *International Journal of Energy Economics and Policy*, (8(3)):216–249.

Tashman, L. (Winter 2018). Beware of standard prediction intervals for causal models. *Foresight: The International Journal of Applied Forecasting*, 48:43–48.

Theocharis, Z. and Harvey, N. (2019). When does more mean worse? Accuracy of judgmental forecasting is nonlinearly related to length of data series. *Omega*, 87:10–19.

Tongal, H. and Berndtsson, R. (2017). Impact of complexity on daily and multi-step forecasting of streamflow with chaotic, stochastic, and black-box models. *Stochastic Environmental Research and Risk Assessment*, 31(3):661–682.

Villar, J. A. and Joutz, F. L. (2006). The relationship between crude oil and natural gas prices. http://aceer.uprm.edu/pdfs/CrudeOil_NaturalGas.pdf Last accessed 2021-04-13.

Wang, Z., Li, Y., Feng, Z., and Wen, K. (2019). Natural gas consumption forecasting model based on coal-to-gas project in China. *Global Energy Interconnection*, 2(5):429–435.

Wearden, G. (2019). Oil prices spike after Saudi drone attack causes biggest disruption ever – as it happened. https://www.theguardian.com/business/live/2019/sep/16/oil-price-saudi-arabia-iran-drone-markets-ftse-pound-brexit\business-live Last accessed 2021-04-13.

Wei, N., Li, C., Peng, X., Li, Y., and Zeng, F. (2019). Daily natural gas consumption forecasting via the application of a novel hybrid model. *Applied Energy*, 250:358–368.

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356.

Wiener, N. (2013). *Cybernetics or control and communication in the animal and the machine*. MIT Press, Cambridge, Mass., 2. ed. edition. ISBN 9781614275022.

Winebrake, J. J. and Sakva, D. (2006). An evaluation of errors in US energy forecasts: 1982–2003. *Energy Policy*, 34(18):3475–3483.

Yalcinoz, T. and Eminoglu, U. (2005). Short term and medium term power distribution load forecasting by neural networks. *Energy Conversion and Management*, 46:1393–1405.

Yaniv, I. and Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4):424–432.

Yu, F. and Xu, X. (2014). A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Applied Energy*, 134:102–113. https://doi.org/10.1016/j.apenergy.2014.07.104.

Zambella, D. and Grassberger, P. (1988). Complexity of forecasting in a class of simple models. *Complex Syst.*, 2(3).

Zellner, A., Keuzenkamp, H. A., and McAleer, M. (2002). *Simplicity, inference and modeling: Keeping it sophisticatedly simple*. Cambridge University Press, Cambridge and New York. ISBN 0521803616.

Zhang, W. and Yang, J. (2015). Forecasting natural gas consumption in China by Bayesian Model Averaging. *Energy Reports*, 1:216–220. https://doi.org/10.1016/j.egyr.2015.11.001.

Čeperić, E., Žiković, S., and Čeperić, V. (2017). Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy*, 140:893–900.

Šebalj, Dario and Dujak Davor and Mesaric Josip (2017). Predicting natural gas consumption - a literature review. http://archive.ceciis.foi.hr/app/public/conferences/2017/08/SPDM-2.pdf.

Šumskis, V. and Giedraitis, V. (2015). Economic implications of energy security in the short run. *Ekonomika*, 94(3):119. https://doi.org/10.15388/Ekon.2015.3.8791.