

# PRODUCTION PLANNING OF ENERGY SYSTEMS

COST AND RISK ASSESSMENT  
FOR DISTRICT HEATING

MAGNUS DAHL

PHD DISSERTATION  
DEPARTMENT OF ENGINEERING  
AARHUS UNIVERSITY, 2018



ISBN: 978-87-7507-422-8  
DOI: 10.7146/aui.270.188

# Contents

Résumé . . . . .	iii
Preface . . . . .	v
Acknowledgements . . . . .	ix
<b>1 Introduction</b>	<b>1</b>
1.1 District heating: past, present and future . . . . .	2
1.2 Risk assessment on different horizons . . . . .	5
1.3 Heat load forecasting . . . . .	7
1.4 About this dissertation . . . . .	9
<b>2 The Aarhus district heating system</b>	<b>11</b>
2.1 District heating in Aarhus . . . . .	11
2.2 The heat load data . . . . .	13
2.3 Heat load and the weather . . . . .	17
<b>3 Weather-based forecast uncertainties</b>	<b>23</b>
3.1 Motivation . . . . .	23
3.2 Methods . . . . .	24
3.3 Main findings . . . . .	26
3.4 Using ensemble weather predictions in district heating operation and load forecasting . . . . .	27
<b>4 The human factor</b>	<b>39</b>
4.1 Motivation . . . . .	39
4.2 Methods . . . . .	39
4.3 Main findings . . . . .	41
4.4 Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data . . . . .	42
<b>5 Seasonal production planning</b>	<b>59</b>
5.1 Motivation . . . . .	59
5.2 Methods . . . . .	59
5.3 Main findings . . . . .	62
5.4 Decision rules for economic summer-shutdown of production units in large district heating systems . . . . .	64
<b>6 Production system planning</b>	<b>75</b>
6.1 Motivation . . . . .	75
6.2 Methods . . . . .	75

6.3	Main findings . . . . .	78
6.4	Cost sensitivity of optimal sector-coupled district heating production systems . . . . .	79
<b>7</b>	<b>Conclusion</b>	<b>95</b>
	<b>Bibliography</b>	<b>97</b>

## Résumé

This dissertation is a collection of research articles that assess economic and operational risk in production planning of district heating. District heating systems are typically coupled to the electricity system through cogeneration and power-to-heat technologies, and production planners must account for uncertainty stemming from changing weather, demands and prices. Years of high-resolution data from the district heating system in Aarhus, Denmark have been used throughout the project to model the system and estimate uncertainties. Risk management tools have been developed to aid district heating operators and investment decision makers in short-, medium- and long-term production planning.

Short-term production planning involves commitment of production units and trading on the electricity markets and relies on forecasts of the heat load. Weather predictions are a significant source of uncertainty for heat load forecasts, because the heat load is highly weather-dependent. I introduce the method of ensemble weather predictions from meteorology to heat load forecasting and create a probabilistic load forecast to estimate the weather-based uncertainty. Better estimates of the weather-based uncertainty can be applied to optimize supply temperature control and reduce heat losses without compromising security of supply in heat distribution systems.

Consumer behavior is another substantial, but difficult to capture, source of uncertainty in short-term heat load forecasts. I include local holiday data in state-of-the-art load forecasts to improve accuracy and capture how load patterns change depending on the behavior of the consumers. A small overall improvement in forecast accuracy is observed. The improvement is more significant on holidays and special occasions that are difficult to forecast accurately.

In medium-term production planning, there can be substantial economic potential in performing summer shutdown of certain production units. The shutdown decision carries significant risk, due to changing seasonal weather patterns. Based on 38 years of weather data, the uncertainty on the timing of the optimal decision is estimated. This information is used to develop practical decision rules that are robust to rare weather events and capable of realizing more than 90% of the potential savings from summer shutdown.

Long-term production planning decisions regarding investments in future district heating production systems are affected by uncertainty from changing electricity prices, fuel prices and investment cost for technology. The effects of these uncertainties on a cost-optimal heat production system are explored, using well-established production and storage technologies and extensive multivariate sensitivity analysis. The optimal technology choices are highly stable and, taxes aside, large heat pumps and heat storages dominate the cost-optimal heat production systems. However, the uncertainty on the exact capacity allocation is substantial. Excluding heat production based on fossil fuels increases the uncertainty on the system cost, but drastically reduces the uncertainty on the optimal capacity allocation.

## Resume

Denne afhandling er en samling af forskningsartikler om økonomisk og operationel risikovurdering i forbindelse med produktionsplanlægning i fjernvarmesystemer. Fjernvarmesystemer er typisk koblet til elsystemet via kraftvarmeproduktion og el-til-varme-teknologier, og produktionsplanlæggere er udsat for usikkerheder fra skiftende vejr, varmebehov og priser. I projektet anvendes adskillige års højtopløst data fra Aarhus' fjernvarmesystem til at modellere systemet og estimere usikkerheder. En række værktøjer til risikohåndtering er blevet udviklet til at hjælpe fjernvarmeoperatører og energiplanlæggere med kort-, mellemlang- og langsigtet produktionsplanlægning.

Produktionsplanlægning på kort sigt indebærer daglig lastfordeling og handel på elmarkederne og er afhængig af varmelastprognoser. En væsentlig kilde til usikkerhed i varmelastprognoser er vejrudsigter, idet varmelasten afhænger af vejret. Jeg introducerer den meteorologiske metode ensemble-prognoser til brug i varmelastprognoser og estimerer den vejrafhængige usikkerhed ved hjælp af en ensemble-varmelastprognose. Bedre estimater af den vejrafhængige prognoseusikkerhed kan anvendes til at optimere styringen af fremløbstemperaturen i varmedistributionssystemer og mindske varmetab uden at gå på kompromis med forsyningssikkerheden.

Forbrugeradfærd er en anden væsentlig kilde til usikkerhed i kortsigtede fjernvarmeprogner, men svær at modellere. For at forbedre prognosepræcisionen og bedre afspejle ændringer i varmelastmønstre afhængigt af varmeforbrugernes opførsel, inkluderer jeg data om lokale ferier og helligdage i varmelastprognoser. Det resulterer i en lille generel forbedring i prognosepræcision. Forbedringerne i prognosepræcision er mere markante i skoleferier og på helligdage, der typisk er særligt svære at prognosticere præcist.

I forbindelse med produktionsplanlægning på mellemlangt sigt, kan der være stort økonomisk potential i at lukke visse produktionsenheder ned sommeren over. Beslutningen om sommerlukning indebærer en væsentlig økonomisk risiko på grund af sæsonbetonede skift i vejret. På baggrund af 38 års vejrdato estimeres usikkerheden på den optimale lukkedato. Resultaterne af denne analyse benyttes til at designe praktiske beslutningsregler, der er robuste over for sjældne vejrbegebenheder og i stand til at realisere mere end 90% af de potentielle besparelser i forbindelse med sommerlukning.

Langsigtede beslutninger angående investeringer i fremtidige fjernvarme-produktionssystemer bliver påvirket af usikkerheder fra ændringer i elpriser, brændselspriser og investeringsomkostninger for teknologi. Konsekvenserne af disse usikkerheder for et omkostningsoptimeret varmeproduktionssystem undersøges ved hjælp af veletablerede produktions- og lagerteknologier og en omfattende multivariat følsomhedsanalyse. Det optimale valg af teknologi er yderst stabilt og domineres af store varmepumper og varmelagre, hvis man ser bort fra skatter og afgifter. Der er dog en væsentlig usikkerhed på den nøjagtige fordeling af produktionskapacitet. Ekskluderes varmeproduktion fra fossile brændsler, så stiger usikkerheden på de totale systemomkostninger, men usikkerheden på den optimale kapacitetsfordeling falder drastisk.

# Preface

This dissertation is the result of three years' work in a research collaboration between Aarhus University and AffaldVarme Aarhus. AffaldVarme Aarhus is the municipal district heating company in Aarhus, and it is where I have spent about 80 % of my time during this industrial PhD fellowship.

Doing research in production planning of district heating, while embedded in the day-to-day operations of a district heating company has been both invaluable and highly challenging. I have been given a rare chance to perform a reality check on my research, both in grand concepts and ideas and in practical complications that may arise at the implementation stage.

Being a part of both the business development department (Forretningsudvikling) and the operations department (Drift) at AffaldVarme Aarhus, I have had a chance to closely follow how a modern district heating company works with strategic long-term production planning and optimization of the daily operations.

Balancing the rigor, reproducibility and novelty required in academic research against the practical needs of an operational organisation has been challenging at times. But I am convinced that this potential conflict has qualified my academic output and made the research more applicable while also benefiting the industrial contribution of my work.

## Industrial contribution

Although the majority of my work these past years has been academic, I have delivered some lasting contributions to AffaldVarme Aarhus.

In 2017 a new straw-fired combined heat and power (CHP) plant opened in Lisbjerg north of Aarhus. This plant significantly changed the operation of the production system, especially during summer. I assisted the engineers in the Operations department by programming a number of heat storage simulations to assess the need for storage in this new configuration of the production system. This work also brought my attention to another interesting production planning issue and research topic: summer shutdown of production units. Chapter 5 covers this extensively and includes a research article I had published in *Applied Energy* about managing risk when shutting down CHP plants over the summer.

My most significant and lasting industrial contribution was the development of an automatic heat load forecasting system. The system predicts the hourly heat load with a horizon of 46 h, based on weather forecasts and data from the heat production system. A machine learning model called support vector

regression performs the load prediction and its performance is competitive with commercial alternatives.

The final version of the system produces an updated forecast every 15 minutes and runs on a server controlled by AffaldVarme Aarhus. My load forecast is intended to supplement a commercial production planning system and assist the system operators in the daily production planning.

I entered a similar load forecasting system into a competition hosted by another large Danish district heating company, Fjernvarme Fyn. My load forecast performed very well, both in accuracy and reliability and won second place in the competition against six other commercial load forecast providers.

## Academic contribution

The remainder of this dissertation is dedicated to my academic output, the core of which is four academic journal articles. I am the first author of these four articles, and they have a chapter each in the dissertation. In each chapter, the main research questions, methods and results are outlined and the research is related to the rest of the project. Besides publishing journal articles I have also presented my research at conferences and written popular outreach articles. A short summary of my academic output follows.

## Journal articles

In order of appearance in the dissertation:

- **Using ensemble weather predictions in district heating operation and load forecasting** by Magnus Dahl, Adam Brun and Gorm B. Andresen. Published in *Applied Energy* (Dahl et al., 2017b).
- **Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data** by Magnus Dahl, Adam Brun, Oliver S. Kirsebom and Gorm B. Andresen. Published in *Energies* (Dahl et al., 2018).
- **Decision rules for economic summer-shutdown of production units in large district heating systems** by Magnus Dahl, Adam Brun and Gorm B. Andresen. Published in *Applied Energy* (Dahl et al., 2017a).
- **Cost sensitivity of optimal sector-coupled district heating production systems** by Magnus Dahl, Adam Brun and Gorm B. Andresen. Submitted to *Energy*.

For all of these articles, I have performed all calculations, mathematical derivations and coding, produced all illustrations and written more than 95 % of the text. No part of this dissertation has previously been considered in preparation for obtaining an academic degree.

## Conference presentations

- **Applications of a heat load forecast with dynamic uncertainties.** Presented at the 2nd International Conference on Smart Energy Systems and 4th Generation District Heating. 2016 Aalborg, Denmark. This presentation won the prize for *Best PhD presentation*.
- **Long-term production planning in large district heating systems.** Presented at the 3rd International Conference on Smart Energy Systems and 4th Generation District Heating. 2017 Copenhagen, Denmark.
- **Machine learning techniques for district heating load forecasting incorporating human behaviour.** Proceeding presented at the 12th Conference on Sustainable Development of Energy, Water and Environment Systems. 2017 Dubrovnik, Croatia.

## Public outreach

I have communicated my research in Danish and in English in magazines that are widely read in the district heating sector, and contributed to the Profile of the Department of Engineering at Aarhus University:

- **Towards realistic production planning** by Magnus Dahl. Published in *Hot Cool – International Magazine on District Heating and Cooling*, no. 1 2017.
- **Realisér værdien i at lukke anlæg ned om sommeren.** Published in *Fjernvarmen, Dansk Fjernvarme's magazine*, no. 1 2018.
- **Forecasts for cheap and sustainable district heating.** Profile 2016, Department of Engineering, Aarhus University.

## Other work

Finally, I have published a book chapter based on my master's thesis, co-authored a journal article and written a few internal technical reports and briefs for AffaldVarme Aarhus. None of these manuscripts are a part of the dissertation.

- **Infrastructure Estimates for a Highly Renewable Global Electricity Grid** by Magnus Dahl, Rolando A. Rodriguez, Anders A. Søndergaard, Timo Zeyer, Gorm B. Andresen and Martin Greiner. Published in the book *New Horizons in Fundamental Physics* (Dahl et al., 2017).
- **Multi-criteria analysis of storages integration and operation solutions into the district heating network of Aarhus – a Simulation Case Study** by Charlotte Marguerite, Gorm B. Andresen and Magnus Dahl. Published in *Energy* (Marguerite et al., 2018).
- **Simulering af sommerdrift med BKVV, SSV og akkumulator-tanke.** March 2017, AffaldVarme Aarhus.

- **Værdien af SSV-VAK i sommerdrift uden Blok 3.** June 2017, AffaldVarme Aarhus.
- **Dokumentation for online varmeprognoosesystem.** March 2018. AffaldVarme Aarhus.

## Funding

This PhD project is part of the READY project (Resource Efficient cities implementing ADvanced smart CitY solutions) which is partly financed by the EU's Research and Innovation funding program FP7 ([https://ec.europa.eu/research/fp7/index\\_en.cfm](https://ec.europa.eu/research/fp7/index_en.cfm)).

## Acknowledgements

I would like to thank Gorm B. Andresen for his scientific vision and excellent supervision. Thanks to him, Adam Brun and Steffen Petersen for making this project possible in the first place.

Thanks to all my colleagues at AffaldVarme Aarhus for making it a great place to work. Thanks in particular to Allan Lundfald, Kim Therkildsen, Grethe Føns Hjortbak and Jeanette Thøgersen for the insight they provided and for teaching me everything I know about district heating.

Thanks to my fellow PhD students at Aarhus University's Department of Engineering for sharing the ups and downs of PhD life. Special thanks to Smail Kozarcanin, Bo Tranberg, Rasmus Høst Pedersen and Hailiang Liu for sparring about the project and giving helpful feedback on manuscripts and presentations.

Thanks to Tom Brown for his part in developing PyPSA and for fruitful discussions about energy systems modeling.

Finally, thanks to my friends and family for believing in me and cheering for me in good times and bad. Especially thanks to my wife Kate Dahl for giving me strength when I needed it and for her unrelenting love and support.



# Chapter 1

## Introduction

The threat of global climate change due to anthropogenic emissions of greenhouse gasses has created a pressure on politicians and energy planners worldwide to transform our energy systems away from fossil fuels. In the European electricity sector, great strides are already being made with large installations of wind and solar power generation, and from 2010 to 2017 electricity generation from fossil fuels was reduced by 17% (Jones et al., 2018). But, there is still a long way to go if any hope is to be retained of limiting global warming to 2°C by the end of the century. Decarbonizing energy systems is a good starting point, since electricity and heat production account for 25% of global greenhouse gas emissions. However, it is crucial that other economic sectors follow suit, particularly the transportation sector and agriculture & forestry, as these sectors are responsible for 14% and 24% of global greenhouse gas emissions (Intergovernmental Panel on Climate Change, 2015).

The electricity sector is currently being decarbonized, primarily through the installation of wind, solar and biomass generation. Wind and solar power generation are variable renewable energy sources, which means that the generation depends on the weather and is not directly controllable by operators. Therefore, electricity systems with large shares of wind and solar generation need a certain amount of dispatchable backup generation, e.g. combined heat and power (CHP) generation. Balancing the supply and demand in a highly renewable electricity system can also be done with large electricity storages, but the technology is not yet mature and cost-competitive on the necessary scale (Schmidt et al., 2017).

The heating sector in the European Union (EU) is responsible for about 40% of the primary energy consumption in the electricity and heating sector (Connolly et al., 2014). Decarbonizing the heating sector and facilitating synergies between renewable power generation and district heating should therefore be a high priority in the planning of future energy systems.

The title of this dissertation is *Production planning of energy systems – cost and risk assessment for district heating*. District heating systems are the topic of the project, and the main focus is planning the production of heat on horizons from a few days to several years. Closer coupling between the electricity and the heating sector introduces new uncertainties for production planners, and planning decisions can carry severe economic and operational risk. In this

project, I not only estimate the cost of a given system or operational scheme, I estimate the uncertainty on the cost and provide tools that production planners can use to quantify and manage operational and economic risk.

## 1.1 District heating: past, present and future

In (Lund et al., 2014), the authors argue that district heating has an important role to play in future energy systems and provide an overview of the history of district heating and a vision for its future. District heating systems will have to be developed further to play a role in highly renewable energy systems of the future, and Lund et al. introduce the concept *4th Generation District Heating* to specify how.

The first generation of district heating used steam as the heat carrier and the first systems were built in the USA in the 1880s. Almost all district heating systems built in Europe and the USA until the 1930s used this technology. District heating was introduced in many cities to replace individual boilers and reduce the risk of boiler explosions. As such, district heating was introduced to provide consumers with safe and convenient heating.

The second generation of district heating used pressurized hot water above 100 °C as the heat carrier. Pipes were often constructed *in situ* and the grid and valves were material heavy. This technology dominated district heating systems built between the 1930s and 1970s, allowed distribution grids to grow larger and improved energy efficiency.

In the third generation, sometimes referred to as *Scandinavian district heating technology*, supply temperatures were decreased to below 100 °C. Pipes for the grid were increasingly prefabricated and pre-insulated, and valve and grid technologies became more material lean. Following the oil crises of the 1970s, oil was replaced with other fuels to a large extent. Third generation district heating technology is predominant in modern district heating systems.

Following the historical overview, Lund et al. proceed to outline the challenges that future district heating systems are faced with in the transition towards more sustainable future energy systems. A fourth generation of district heating will need to operate at lower temperatures. Lower temperatures reduce heat losses from the grid and facilitate better integration of renewable energy sources and industrial waste heat. In addition, the electricity consumption for heat pumps can be drastically reduced by decreasing supply temperatures.

In chapter Chapter 3, I present a scheme for temperature control in heat distribution grids. The control scheme is designed to reduce the supply temperature, without compromising the security of supply. This is achieved by linking the temperature control to a probabilistic heat load forecast, allowing for more aggressive temperature control when the load forecast is very certain. My control scheme is not as ambitious with respect to temperature reductions as the very low temperature grids proposed in (Lund et al., 2014). However, the purpose of the control scheme is the same: to improve energy efficiency by reducing heat losses from the grid.

Besides very low grid temperatures, the fourth generation of district heating must also be able to integrate with a low-energy building environment. Building

modeling and demand side control is beyond the scope of this project, but there is a large potential in combining building retrofit with expansions of district heating grids. Combining lower supply temperatures and lower residential heating demands it becomes possible to supply heat to more consumers using the same pipes. This is especially relevant in urban areas with growing population density, because expanding grid infrastructure is quite costly (Frederiksen and Werner, 2013).

Finally, it is imperative that future district heating systems integrate well with the electricity sector (Lund et al., 2014). CHP generation can provide backup generation for a highly renewable electricity system, but lower electricity prices due to surplus wind generation risk undermining the economy of CHP plants (Traber and Kemfert, 2011).

### **Large-scale power-to-heat**

Power-to-heat technologies such as electric boilers and heat pumps can be used to absorb surplus electricity generation into heating systems, and provide much needed flexibility for the electricity sector.

Large-scale deployment of power-to-heat technologies is not unrealistic, as demonstrated in (Averfalk et al., 2017), where the authors review the operational experience with large heat pumps for district heating in Sweden. Since the 1980s, Sweden has installed the world's biggest fleet of heat pumps and electric boilers. Initially, this power-to-heat capacity was built to absorb a national surplus of electricity from nuclear power that could not be exported due to weak international transmission lines. Most of these heat pumps are still in operation, and most of the heat is supplied from sewage water and ambient water, i.e. sea water, lake and river water and ground water. Industrial waste heat constitutes a smaller amount of the heat supply, and its capacity utilization has gone down significantly since the 1980s. Averfalk et al. conclude that sewage water and ambient water are more stable long-term heat sources for heat pumps compared to industrial waste heat.

Large heat pumps still play an important role in Swedish district heating systems, but their capacity utilization has decreased somewhat, because of increased competition from biomass and waste incineration CHP. The heat supplied from power-to-heat technologies in Swedish district heating systems peaked in 1990. Taxes on electricity also has potential to decrease the competitiveness of power-to-heat technologies.

Besides taxes on electricity, there are other potential technical and environmental roadblocks for large-scale implementation of heat pumps in district heating systems globally. For many years, chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) were used as refrigerants in large heat pumps. Due to their depleting effect on the ozone layer CFCs and HCFCs were banned as refrigerant in the Montreal Protocol in 1989. In the large Swedish heat pumps, these refrigerants have been replaced with hydrofluorocarbons (HFCs), which do not deplete the ozone layer of the atmosphere. However, HFCs are extremely potent greenhouse gasses, and may contribute to global warming. The use of HFCs is therefore discouraged or banned for large-scale applications in many countries. Natural refrigerants do exist as alternatives, e.g. hydrocarbons,

ammonia or carbon dioxide. These alternative refrigerants each have their issues, but Averfalk et al. expect HFCs to be replaced by natural refrigerants in the future.

Despite potential problems, power-to-heat technologies, especially large heat pumps, have a significant role to play in the future's district heating systems. They are a robust and well-established technology, with potential to provide flexibility to the electricity system and provide heat at a competitive cost. It is emphasized that power-to-heat technologies are best used in energy systems with continuous surplus of electricity generation. This surplus can stem from variable wind and solar power, or from base load nuclear and hydro power (Averfalk et al., 2017). Periodical electricity surplus is exactly the state of the current and near future European electricity system, due to large expansions in wind and solar power generation (Jones et al., 2018).

In chapter Chapter 6, I propose a cost-optimal heat production system for a city. Power-to-heat technologies, specifically large heat pumps, play a very important role in the cost-optimal system, regardless of whether or not a fossil free policy is adopted. Large heat storages help the heat pumps provide the necessary heat while maintaining a good economy with respect to the electricity market. Taxes and tariffs aside, I recommend massive investments in large heat pumps and heat storages, when constructing new production infrastructure.

### **The future potential of district heating**

As mentioned, the heating sector is responsible for about 40% of the total primary energy consumption for heat and electricity in the EU. District heating only has a market share of 13% of the heating sector. However, it is demonstrated in (Connolly et al., 2014) that there is significant potential for expanding the use of district heating in the EU.

In the EU's heating and electricity system, almost half of the primary energy is lost before being delivered to the end user (Connolly et al., 2014). Most of this loss occurs in the conversion to electricity, and district heating systems can utilize much of the wasted energy through CHP generation. Although losses in heating grids are not negligible, expanding the coverage of district heating and integrating the heating and electricity sectors further can dramatically increase the energy efficiency of the whole energy system. This is what makes district heating economically competitive.

Connolly et al. map out the district heating potential in the EU using geographic information systems (GIS) combined with energy system modeling. The geographical density of heat demand is estimated from national statistics and population densities with a resolution of 1 km<sup>2</sup>. District heating is only economically viable in areas with high population density, because heating grids require large investments, and heat losses become too large in dispersed grids with low heat consumption.

The proposal for the future EU heating sector put forward in (Connolly et al., 2014) is compared to a reference scenario proposed by the European Commission in the report *Energy Roadmap 2050* (European Commission, 2011). The European Commission's scenario reduces greenhouse gas emissions by 80% compared the 1990 levels through a combination of energy savings and electrification of the

heating sector. Connolly et al. demonstrate that implementing the necessary large-scale savings on electricity and heat will not be cost-effective. Instead, they propose massive expansion of district heating in densely populated areas as a means to improve the overall energy efficiency of the electricity and heating sector. The district heating market share is increased to 30 % in 2030 and to 50 % in 2050, and the proposed solution meets the same ambitious emission target, but is approximately 15 % cheaper (Connolly et al., 2014).

In rural areas, where the heat demand density is insufficient for district heating, individual heat pumps are suggested as an alternative. Large-scale heat pumps for district heating have several advantages over individual heat pumps. Large-scale heat pumps benefit from the economy of size and can be used to balance the electricity grid, due to the central control and large capacity (Averfalk et al., 2017). Individual heat pumps, however, do have a role to play in the future heating sector in rural areas, especially as a way to limit the consumption of biomass in a 100 % renewable energy system (Mathiesen et al., 2012).

Summing up, district heating is an important player in future renewable energy systems. While district heating systems have come a long way during the last century, there are still improvements that can be done. This dissertation is focused on improving operation and production planning through better modeling of uncertainties and risks.

## 1.2 Risk assessment on different horizons

The International Organization for Standardization defines risk as "the effect of uncertainty on objectives" (ISO, 2009). District heating providers face many different uncertainties in the daily operation and production planning. Relevant sources of uncertainty include: the weather, consumer behavior, changing heat and electricity demands as well as the near and far future cost of electricity, fuels and relevant technologies.

The *objectives*, that these uncertainties affect, fall mainly in two categories. The first objective of most modern district heating providers is to always deliver heat to cover the consumer demand, i.e. security of supply. The second objective is to minimize the total cost of producing and distributing heat, without compromising security of supply. Often, there are other tertiary or competing objectives, most notably environmental protection and climate goals, but the risk assessment in this project is primarily focused on the two main objectives. The effect of uncertainties on security of supply I call *operational risk*, and the effect of uncertainties on operational or investment cost I call *economic risk*.

Risk assessment of production planning in district heating systems can be divided according to the planning horizon into short-term, medium-term and long-term planning.

Short-term production planning, on horizons up to a few days, primarily deals with commitment of production units and planning of the operation of heat storage tanks. Short-term planning in systems with cogeneration or power-

to-heat technologies must account for changing electricity prices as well as the heat demand in the days to come.

Medium-term or seasonal production planning takes place on horizons up to several months. Decisions, concerning e.g. complete shutdown of larger production units for maintenance or economic reasons, fall in this category and are influenced by large seasonal variations in the heat load.

Long-term production planning, on horizons of several years, deals with investment decisions. Sizing of new production systems and expansion of existing systems requires planning with long time horizons. Long-term planning decisions are impacted by changing demands of heat and electricity, variations in electricity and fuel prices and in the cost of relevant technologies.

In (Carpaneto et al., 2011a), the authors address short- and medium-term planning in cogeneration systems under uncertainty. Carpaneto et al. present an ambitious framework for planning of cogeneration on multiple time scales under uncertainty from electricity price variations and changing energy loads. Under three different operational strategies, the authors characterize the operation of a single CHP unit using nonsequential Monte Carlo simulation. The correlations between different random input variables are captured, but the autocorrelations are lost. The nonsequential nature of the framework makes it unsuitable for incorporating storage operation, which is highly relevant when planning the production in larger district heating systems.

In a companion paper to (Carpaneto et al., 2011a), the authors address the problem of long-term planning of cogeneration systems using decision theory (Carpaneto et al., 2011b). Carpaneto et al. argue that the uncertainties on planning horizons as long as 10-20 years are so large that decision theoretical approaches are more suitable for characterizing the variety of possible solutions compared to traditional statistical approaches. Using decision criteria such as minimizing the expected value of a financial objective or minimizing the impact of the worst case scenario, the paper presents more than just one right choice for decision makers. However, such a framework is only useful to decision makers who are very clear on their priorities, values and risk tolerance.

In (DeCarolis, 2011), the author provides an interesting alternative way to handle uncertainty in long-term production planning of energy systems. DeCarolis argues that many energy system modelers address future uncertainties by creating vast elaborate models and running a few scenarios with different input assumptions. This approach is inflexible and provides little insight. DeCarolis proposes a technique from operations research called modeling to generate alternatives (MGA), as way to address future uncertainties. The proposed approach explores near-optimal solutions that are maximally different in decision space. This is a systematic way to generate alternatives and highlight plausible alternatives in the future constitution of the energy system.

The research in this project assesses both economic and operational risk in short- medium- and long-term production planning.

In Chapter 3, a strategy is presented to manage operational risk with respect to control of supply temperatures for area substations. By reducing supply temperatures, heat losses from the grid can be limited, but if the temperature is too low, there is a risk that the demand cannot be met. The proposed control

strategy uses an estimate of the weather-based uncertainty on the forecasted demand to reduce supply the temperature while maintaining security of supply. This is operational risk management on the distribution side of the district heating system. Moreover, the estimated forecast uncertainties are a valuable tool for short-term production planning.

In Chapter 5, the operational risk related to the production side of the district heating system is assessed. It is demonstrated that there are large potential savings associated with shutting down certain production units during the summer. However, the optimal dates for shutdown in the spring and start-up in the fall vary greatly depending on seasonal weather patterns. If a plant is shut down too early and another unit falls out, the production system may be incapable of delivering the necessary heat. This is a weather-based operational risk, that I assess using 38 years of hourly weather data. On this very solid data foundation, I create decision rules that can help production planners make good medium-term planning decisions while managing the operational risk in a cost-effective way.

In addition to operational risk, the decision rules in Chapter 5 were also designed to manage economic risk related to seasonal production planning decisions. Shutting down a plant too early or too late is costly, and the developed decision rules can assist production planners in managing the economic risk related to the decision.

The topic of Chapter 6 is the cost-optimal installation of production capacities in a city-wide district heating system. Long-term investment decisions are impacted by many uncertainties, regarding future demands and prices of fuels, electricity and production technologies. The economic risk related to the investment in new district heating systems is assessed through extensive multivariate sensitivity analysis.

Summarizing, the research in this project covers both operational and economic risk assessment for production planning on time horizons from a few days to several years.

## 1.3 Heat load forecasting

In the daily operation and production planning of modern district heating systems, one of the most important tools for managing operational and economic risk is reliable forecasts of the heat load on the system. The heat production of the following day is planned while accounting for bids in the day-ahead electricity market. Short-term heat load forecasts with horizons up to 48 h are particularly relevant in the daily production planning.

An important part of the research in this dissertation is devoted to improving short-term heat load forecasts and quantifying their uncertainties. I therefore digress slightly in this section and dive into a few of the technicalities of load forecasting. Long-term heat load forecasts, with horizons of several years, are used for sizing of district heating grids and production systems, and are often created based on population growth models. While long-term forecasts are relevant for planning of future systems, my work is focused on short-term heat

load forecasts, and the term *heat load forecast* throughout this dissertation refers to short-term forecasts.

The process of creating a load forecast can be divided into three elements: input, model and output.

The input is the data that is used to train the model and predict the load. Typical input data for heat load forecasts include weather variables such as outside temperature, wind speed and solar irradiance. It is also common to use the time-lagged load as input to forecast models.

The model is a mathematical mapping from input to output. Usually, parameters in the model are adjusted by training the model (fitting), in order to minimize the deviation between the output from the model and the desired output. In the research literature, statistical time series models and machine learning models are most predominant for forecasting load in district heating systems (see e.g. (Kato et al., 2008; Grosswindhager et al., 2011; Idowu et al., 2014; Fang and Lahdelma, 2016)). Bottom up physical models are difficult to build for district heating systems due the high complexity of the system, but knowledge about the physics of the system can be used to improve statistical models (Nielsen and Madsen, 2006). In civil engineering, there is a strong tradition for building energy models (BEMs). These models can achieve accurate predictions of the energy consumption of a single building, but require detailed input about geometry, exact location, materials etc. This data is typically not available for all the dwellings in a district heating system, making BEMs hard to use for district heating load forecasting. There are, however, new advances using Bayesian calibration of BEMs on clusters of buildings, which may be a first step towards more elaborate district load forecasts (Kristensen et al., 2017).

Finally, the last element of a load forecast is the output. The output is simply the prediction of the load, given a model and a set of input data. Comparing the output of the model to the actual heat load is the most common way to measure the performance of the forecast.

Building a good load forecast is an iterative procedure of selecting the input data, selecting, tuning and training the model and evaluating the performance. It is crucial that this procedure is carried out in a structured way to ensure that the forecast has good generalization performance. This means that the forecast model performs well on previously unseen data. A common pitfall is overfitting, a condition in which the model fits known data very well, but performs poorly in out-of-sample predictions. By evaluating forecast performance on blind test data that was not used for training the model, it is possible to get a better estimate of the generalization performance of the forecast. A more sophisticated version of this procedure is cross-validation, which better utilizes the data (Hyndman and Athanasopoulos, 2014). It is important to stress that good forecasting practices, especially regarding performance evaluation, are indispensable in the process of creating high quality load forecasts.

It is difficult to compare performance fairly between different studies on district heating load forecasting. Studies are often system-specific, and there is not a tradition of benchmarking forecast models on a standard dataset as in e.g. image recognition studies. In my work with creating heat load forecasts, I have primarily used weather variables and time-lagged load data as input. I have

observed a performance increase when moving from very simple linear models, to more complex machine learning model such as support vector regression and simple neural networks. However, these models perform similarly on the same input data, and it has not been possible for me to improve performance further, by using more advanced models. Given the same input data, there seems to be a plateau in performance.

I believe that the next big advancement in heat load forecasting performance will come from incorporating new types of input data and not just from employing more sophisticated models. District heating load is driven by consumer demand and being able to more accurately predict the behavior of consumers could lead to better forecast performance. In Chapter 4, I use local holiday data as a proxy to model human behavior in the forecast models. The performance improvements were minor, but not insignificant on special days that are traditionally difficult to forecast accurately (i.e. New Year's Eve). Similar approaches and creative use of new data sources may lead to better heat load forecasts in the future.

Data availability is higher than ever, also in district heating systems. Wireless gathering of heat meter data for billing and control purposes is becoming increasingly available. Hydraulic data that was previously estimated through elaborate mathematical models (e.g. (Madsen et al., 1994; Arvastson, 2001)), can be measured at a reasonable cost today. Weather data, both measurements and forecasts are also becoming easier to obtain. The Norwegian Meteorological Institute already provides free weather data, and the Danish Meteorological Institute has been announced to follow suit between 2019 and 2023.

The improved availability of weather data is good news for heat load forecasters. It is important to remember, that the real test of a load forecast is the performance that can be achieved in actual live operation. An good live system has to produce a load forecast without human interference, even if data feeds are missing or corrupted. Throughout this project, I have implemented several live load forecasting systems, and sometimes practicalities, such as a lack of real-time data, can make a more sophisticated model useless. Adding more data sources may substantially increase the technical debt of the system and make it more vulnerable (Sculley et al., 2015). Reliability is just as important as accuracy when creating tools for the everyday production planning in district heating systems.

## 1.4 About this dissertation

Energy systems are changing rapidly with extensive installation of power generation from variable renewable energy sources. District heating can play an important role in future sustainable energy systems, but the sector must adapt to the new reality. Increased integration between the heating and electricity sectors introduces new uncertainties for district heating planners and makes risk assessment more important than ever.

The body of research presented in this dissertation provides tools and insight relevant to risk assessment for both short-, medium- and long-term production planning of district heating. Chapters 3 and 4 are focused on heat load forecasts, that are valuable tools for short-term production planning. Chapter 5 deals

with medium-term seasonal planning decisions, and Chapter 6 treats economic risk regarding long-term investments in future district heating systems.

Each of the four Chapters 3-6 are built around a research article with a small introduction to the motivation, methods and main findings of the article.

Chapter 2, *The Aarhus district heating system* is a brief introduction to the Aarhus district heating system which served as the primary study case and data provider for the project. Key parts of the project's data foundation are also presented here.

Chapter 3, *Weather-based forecast uncertainties*, addresses the problem of weather-based uncertainties in short-term load forecasts. The technique of ensemble weather prediction is borrowed from meteorology and applied to create a probabilistic load forecast. The weather-based forecast uncertainties are applied to a temperature control scheme, capable of reducing heat losses without compromising security of supply.

Chapter 4, *The human factor*, treats the problem of including the behavior of the heat consumers in load forecasts. The potential of improving heat load forecasts using local holiday data to better capture the human component is assessed. In addition, this chapter quantifies how much load forecasts can be improved by improving weather forecasts.

Chapter 5, *Seasonal production planning*, is about assessing and managing the risk related to performing seasonal shutdown of a large production unit in a district heating system. The uncertainty on the optimal decision stems from variations in seasonal weather patterns from year to year and is estimated using an extensive weather dataset. Decision rules are developed to aid production planners in taking the best decisions in the face of uncertainty.

Finally, Chapter 6, *Production system planning*, addresses the economic risk on cost-optimal investment decisions for future district heating systems. Variations in electricity and fuel prices and developments in investment costs are uncertainties that affect cost-optimal allocation of production capacities. The effects of these uncertainties on cost-optimal heat production systems are estimated by means of extensive multivariate sensitivity analysis.

## Chapter 2

# The Aarhus district heating system

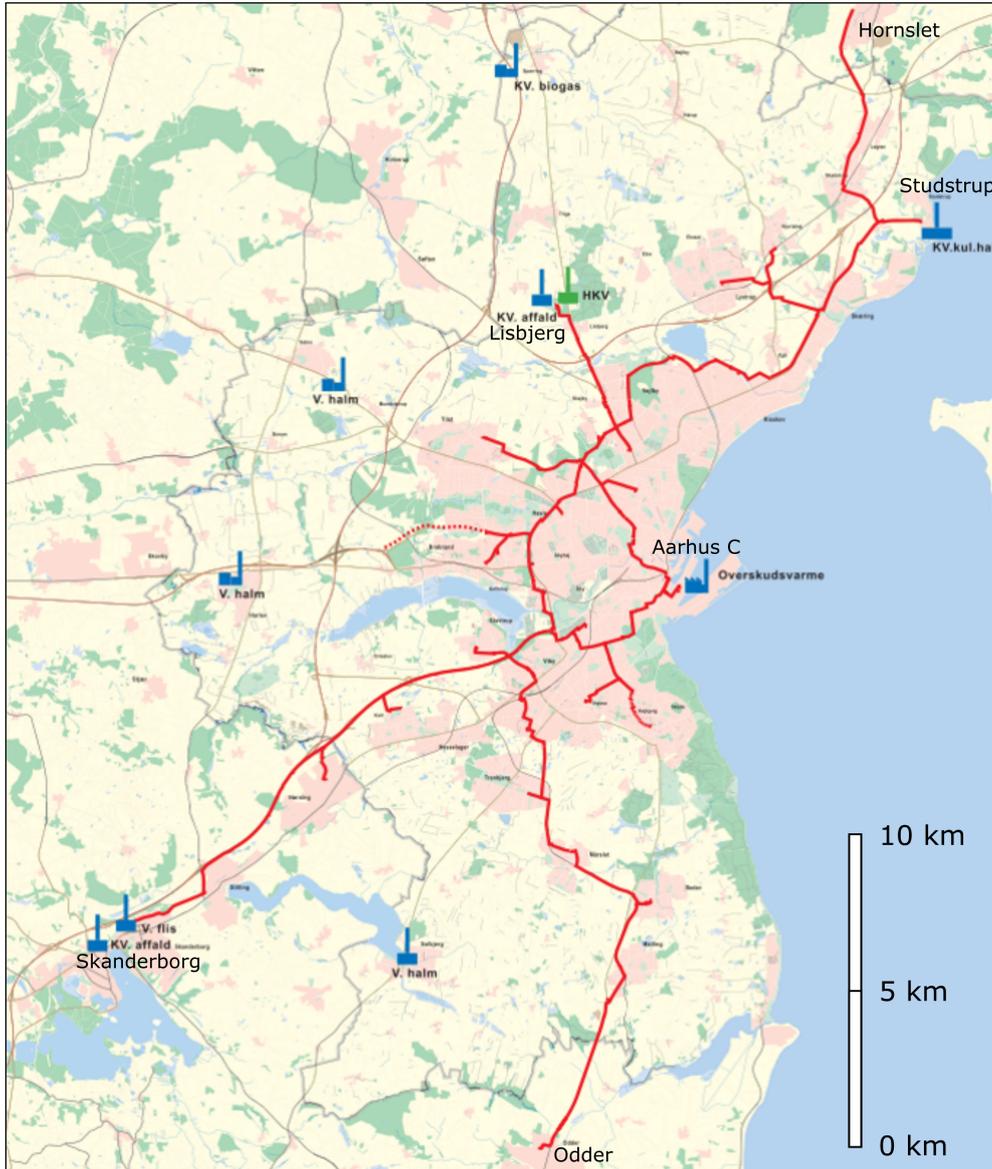
This chapter is intended as a brief introduction to the district heating system of Aarhus, Denmark and a presentation of a key part of the data foundation for the project: seven years of hourly heat load paired with historical weather data. The Aarhus district heating system serves as a case study and data provider throughout the project.

### 2.1 District heating in Aarhus

Aarhus is the second-largest city in Denmark, with a population of about 340,000 people as of January 2018 (Aarhus Kommune, 2018), and about 95 % of the population have district heating.

The seed of the modern Aarhus district heating system was planted in the early 1980s, when it was decided to join a number of local heat distribution grids together by building a large transmission grid, owned by the municipality. Heat production was transformed from local oil boilers to large centralized CHP production fueled by coal. The centralization was motivated by the oil crises in the 1970s and became the foundation of one of the largest district heating systems in Denmark.

Today, the Aarhus district heating system consists of a transmission system and a distribution system. The transmission system receives heat from most of the production units and operates at high supply temperatures with large pipe diameters. Temperatures in the transmission system vary over the year, but typical values are about 105 °C for supply and 45 °C for return temperature. The transmission system spans 38 km from north to south, and consists of more than 136 km pipes. A map of the transmission system as of 2010 is shown in Figure 2.1. The distribution system delivers the heat to the end users and comprises a number of smaller grids. The distribution system is coupled to the transmission system through about 50 area substations with heat exchangers. The distribution grids consist of more than 2,000 km pipes and operate at lower temperatures compared to the transmission grid, typically 75 °C for supply and 45 °C for return temperature.



**Figure 2.1:** The Aarhus transmission system as of 2010 is drawn in red. Some of the heat production units in the system are shown, with *V* denoting heat only boilers and *KV* denoting CHP production. The dotted transmission line was planned at the time and has since been completed.

The division between transmission and distribution grids only exist in the largest few district heating systems in Denmark. A vast majority of Danish district heating systems are smaller and operate without a transmission system, and with fewer production units.

Since the 1980s, the Aarhus heat production system has incorporated an increasing amount of heat from waste incineration CHP, and in 2015 it covered 25% of the annual heat load. A large CHP plant located in Studstrup north of Aarhus has been providing a majority of the heat since the 1980s. The plant, *Studstrupværket*, was expanded in the 1980s and burned coal until 2016, when

it was converted to wood pellets. The primary heat production from coal and waste incineration has been supplemented by oil boilers in peak load situations, and in the last few years, an electric boiler, has been able to provide heat when electricity prices are low. Finally, a number of production units using biomass and industrial waste heat supplement the base load with a small combined capacity. With the completion of a large straw fired CHP plant in Lisbjerg in 2017, Aarhus reached the goal of providing fossil free district heating.

Summing up, the heat production system of Aarhus has been through a transformation from decentralized oil boilers, over centralized coal CHP to biomass CHP and waste incineration. Although the heat production is now entirely fossil free, new challenges arise in the horizon. The large Studstrup CHP plant, now fired by wood pellets, is coming to the end of its lifetime around 2030, and it will be necessary to decide on a replacement. Given the time scale of planning large-scale production infrastructure, this work must begin soon.

Detailed technical and economic data about the heat production units in Aarhus are used for realistic production simulations in Chapters 5 and 6.

## 2.2 The heat load data

Throughout this dissertation, *heat load* refers to the load on the heat production system. At all times, the production system must deliver an amount of heat equal to the consumer demand plus the heat that is lost in the system. The heat loss in Aarhus is approximately 15-20 %, and most of the heat is lost in the distribution grids. I measure the total heat load on the system as the total output from all the heat production units.

All the studies in this project are based on heat load from the Aarhus system. Seven years of hourly heat load data from 2009 to 2016, except 2011, has been provided by AffaldVarme Aarhus, the municipal district heating company. I chose not to include data from 2017, due to many data errors and fallouts in connection with the opening of the new straw fired CHP plant in Lisbjerg.

In Figure 2.2(a-b), the hourly heat load time series for the seven years are plotted. Inspecting the plots, it is clear that although the heat load is different from year to year, there are some common characteristics between the years. Unsurprisingly, the heat load is high in winter and low in summer. The daily variations in the load are also significantly larger in the winter.

A few data errors appear to have been corrected, e.g. in November of 2012, but overall this is a high quality dataset that can provide good insight into the variations in the heat load over several years.

The heat demand in Danish district heating systems can be divided into heat for space heating and heat for hot water consumption. The demand for space heating is the main cause of the difference between summer and winter loads, as it is highly dependent on the weather. The heat demand for hot water consumption changes little over the year, although the water that is heated for consumption may initially be colder during winter. The summer base load is dominated by hot water consumption.

The heat load time series from Figure 2.2(a-b) are summarized in Tables 2.1 and 2.2. The monthly means of the heat load of Aarhus are shown in Table 2.1.

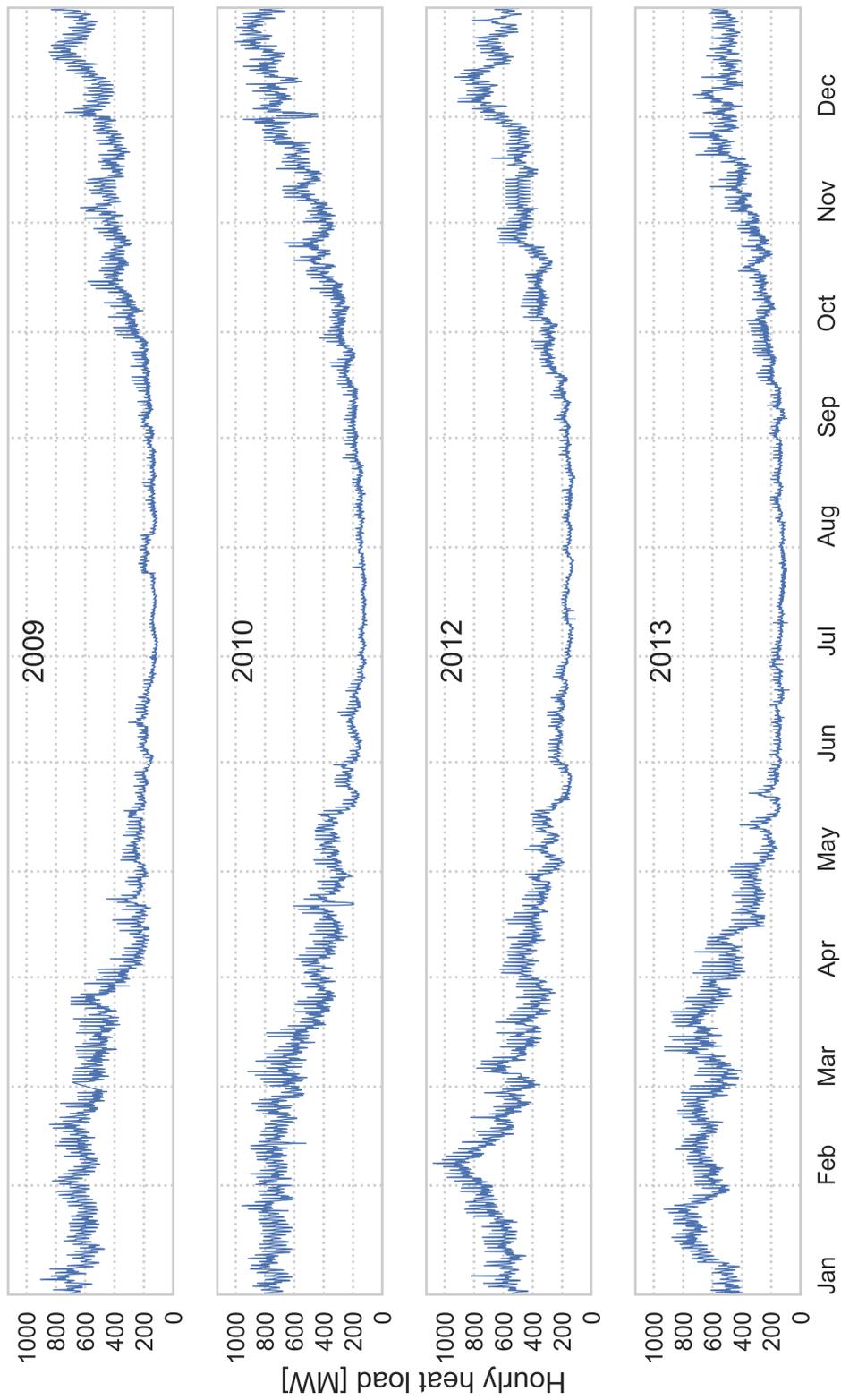


Figure 2.2: (a) Hourly heat load for Aarhus, in 2009, 2010, 2012 and 2013.

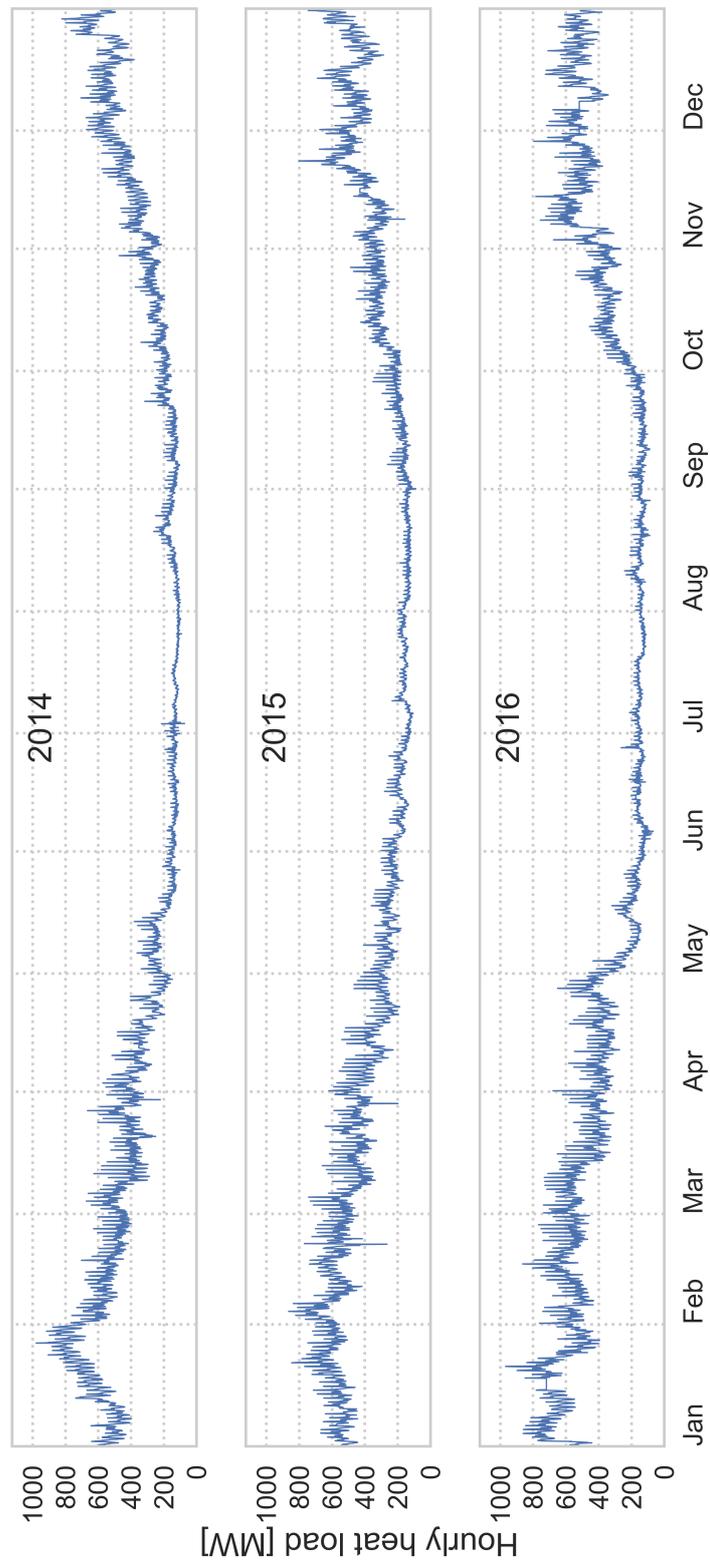


Figure 2.2: (b) Hourly heat load for Aarhus, in 2014, 2015 and 2016.

July and August tend to have the lowest mean loads and January and February tend to have the highest loads. Most of the months with the highest load (in bold) fall in the years 2010 and 2012, and 2010 has the highest annual mean load with 418 MW. This is 95 MW higher than the mean load of 2014, the year with the lowest load. Thus, while there are clear patterns in the monthly mean load, there are also significant variations in the load from year to year.

**Table 2.1:** Estimated mean of the heat load by month. The maximum value within each month is highlighted. The last row shows the mean load over the entire year.

	Mean heat load [MW]						
	2009	2010	2012	2013	2014	2015	2016
January	630	<b>743</b>	621	637	636	578	647
February	617	<b>712</b>	689	656	540	590	576
March	502	533	444	<b>625</b>	430	467	488
April	250	368	401	<b>402</b>	320	345	395
May	226	<b>293</b>	244	213	207	253	201
June	176	173	<b>208</b>	148	137	184	146
July	144	131	<b>155</b>	128	124	155	147
August	144	<b>158</b>	151	139	147	139	146
September	186	223	<b>224</b>	182	155	179	141
October	355	364	<b>374</b>	269	245	301	331
November	433	<b>561</b>	486	450	394	414	513
December	605	<b>772</b>	674	510	558	473	522
Annual	355	<b>418</b>	388	362	323	338	354

**Table 2.2:** Estimated standard deviation of the heat load by month. The maximum value within each month is highlighted. The last row shows the standard deviation of the load estimated for the entire year.

	Standard deviation of heat load [MW]						
	2009	2010	2012	2013	2014	2015	2016
January	71	63	89	124	<b>131</b>	68	106
February	75	73	<b>150</b>	68	76	79	72
March	78	<b>119</b>	100	95	75	72	95
April	58	79	72	<b>103</b>	86	83	63
May	38	<b>75</b>	73	60	57	39	57
June	<b>33</b>	31	30	19	14	32	22
July	<b>29</b>	12	17	16	13	20	18
August	21	22	16	16	<b>31</b>	14	20
September	33	42	<b>57</b>	43	35	36	22
October	66	<b>82</b>	80	48	49	56	66
November	60	<b>123</b>	57	89	84	103	85
December	99	<b>103</b>	91	62	73	80	70
Annual	194	<b>240</b>	204	211	189	168	195

Table 2.2 summarizes the standard deviation of the heat load estimated within each month, a measure of the magnitude of the load variations. The variations tend to be smallest in the summer months and largest the cold months from November to April, although the picture is not as clear as for the mean loads. Again, many of the months with the highest standard deviation fall in 2010, but the highest monthly standard deviation is observed in 2012. February of 2012 displays loads as high as 1,080 MW and as low as 386 MW, resulting in a standard deviation of 150 MW for the month.

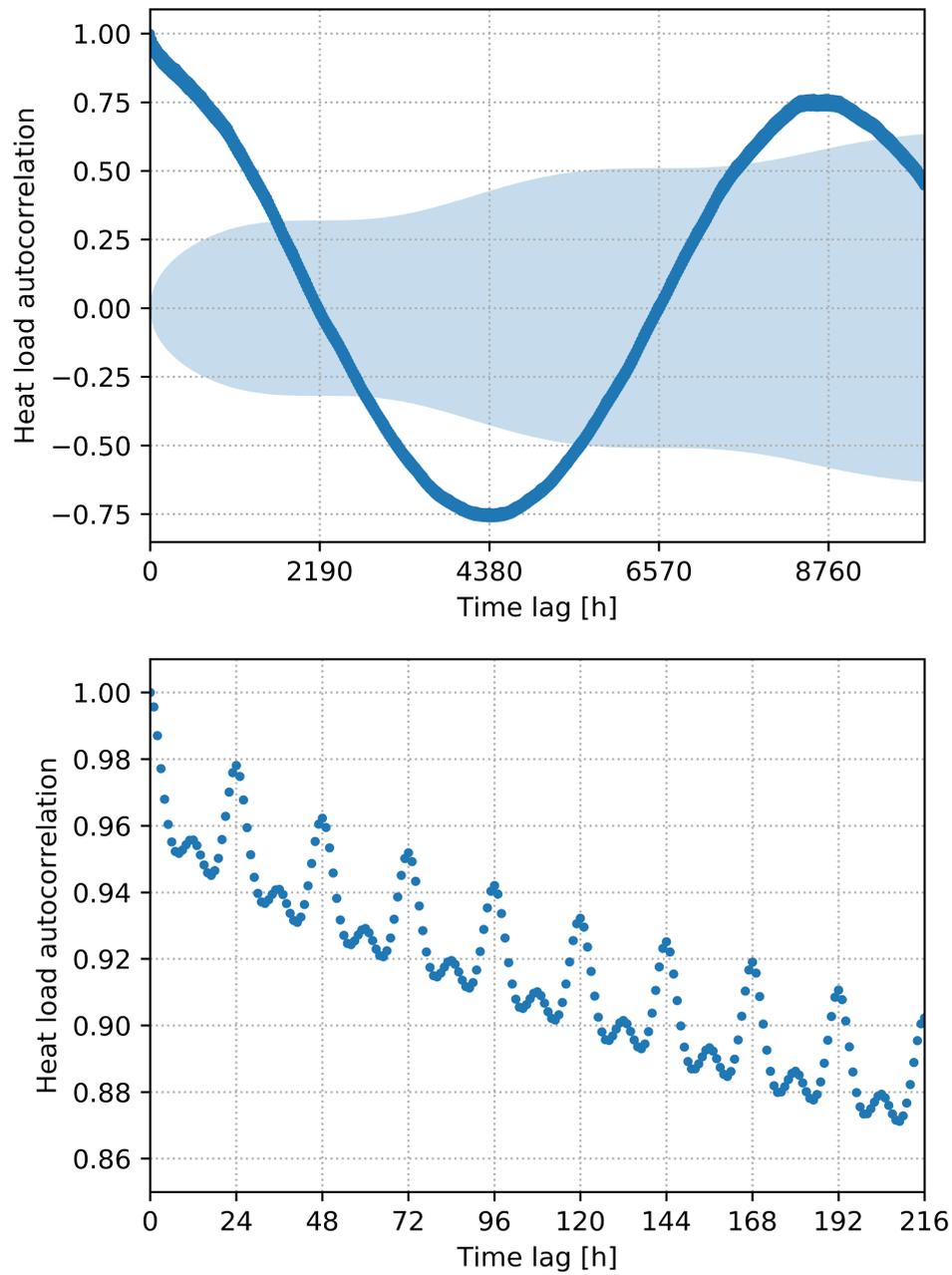
The periodic patterns in the heat load can be explored in several ways, but since forecasting is an important part of this project, the autocorrelation function is used here. The autocorrelation of a time series is the Pearson correlation of the time series with a time-lagged copy of itself, as a function of the time lag. Figure 2.3 shows the autocorrelation function of the heat load, estimated from the seven years of hourly data. The top figure shows the long-term structure of the autocorrelation and illustrates the annual periodicity of the heat load. The time lag is shown in hours and 4,380 h is one half year and 8,760 h is one year. The negative peak at half a year and the positive peak at a whole year indicate a strong annual pattern in the load, a pattern that reflects the annual variations in the weather.

Zooming in, the bottom panel of Figure 2.3 shows the short-term structure in the autocorrelation. There are peaks at every multiple of 24 h, revealing a distinct daily periodicity in the load. The daily pattern in the load stems from the human component in the district heating system: heat consumers who follow a daily schedule and shape the consumption pattern. When setting out to create a forecast of a time series, it is helpful to familiarize oneself with the timewise structure of the data. In Chapters 3 and 4, the strong correlation with loads lagged by 24 h and 48 h is exploited to create heat load forecasts with horizons up to two days.

## 2.3 Heat load and the weather

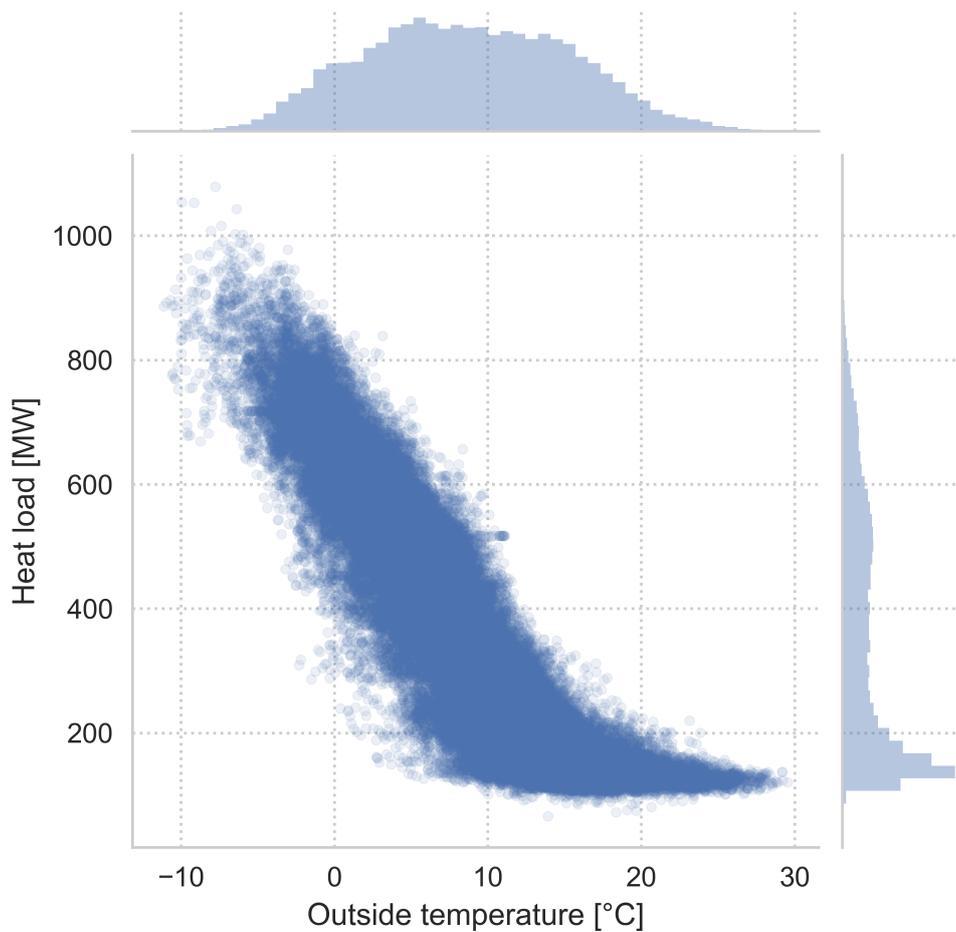
The load in a district heating system is highly dependent on the weather. In Scandinavian heating systems, the load primarily depends on the outside temperature and, to a lesser degree, the wind speed and solar irradiance (Frederiksen and Werner, 2013; Arvastson, 2001).

To illustrate the weather dependence of the heat load, I use data from the National Centers for Environmental Prediction (NCEP) under the United States National Weather Service. Data about air temperature, wind speed and solar radiation is sourced from two high-resolution global reanalysis datasets CFSR and CFSv2 (Saha et al., 2010, 2011). This data is used to train the machine learning forecast models in Chapter 4 and to simulate the heat load in Chapter 5. Where available, measurements from local weather stations may be more accurate, but the quality of the CFSR and CFSv2 data is high, and it is available for all years from 1979 to 2018. Using data from such a long time period can help capture rare weather events and their impact on district heating systems.



**Figure 2.3:** The autocorrelation function for the heat load of Aarhus estimated from seven years of hourly data. The top figure shows the long-term correlations, with 8,760 h constituting one year. The bottom figure is a zoom to show the short-term correlations up to nine days. 95 % confidence intervals are shown in light blue.

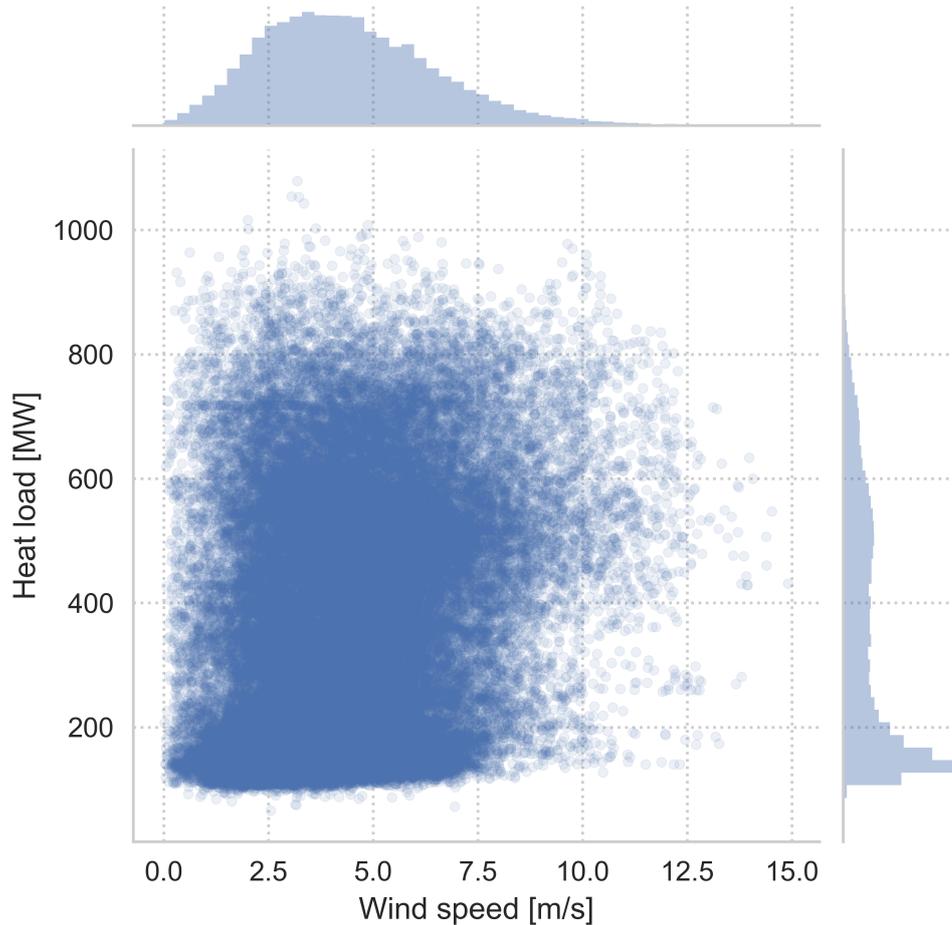
In Figure 2.4, the heat load is plotted against the outside temperature for each hour in the seven years: 2009-2016, excluding 2011. The load versus temperature forms a hockey stick shape, with a linear relationship at low temperatures and a constant load at higher temperatures. The hockey stick shape is well-known in district heating literature and reflects the fact that the temperature dependence of the load subsides as the weather gets warmer (Frederiksen and Werner, 2013). The linear relationship during the cold period is strong, and the Pearson correlation coefficient is  $-0.90$  between the hourly load and the outside temperature.



**Figure 2.4:** Scatter plot of the hourly heat load versus the outside temperature over seven years. The histograms in the side panels illustrate how the variables are distributed. The Pearson correlation coefficient is  $-0.90$  between heat load and outside temperature.

In the side panels of Figure 2.4, histograms illustrate how the heat load and the temperature are distributed individually. While the temperature distribution is fairly symmetric, the heat load distribution is heavily skewed towards low loads, with a long tail of high loads. This is another reflection of the nonlinearity in the load-temperature relationship.

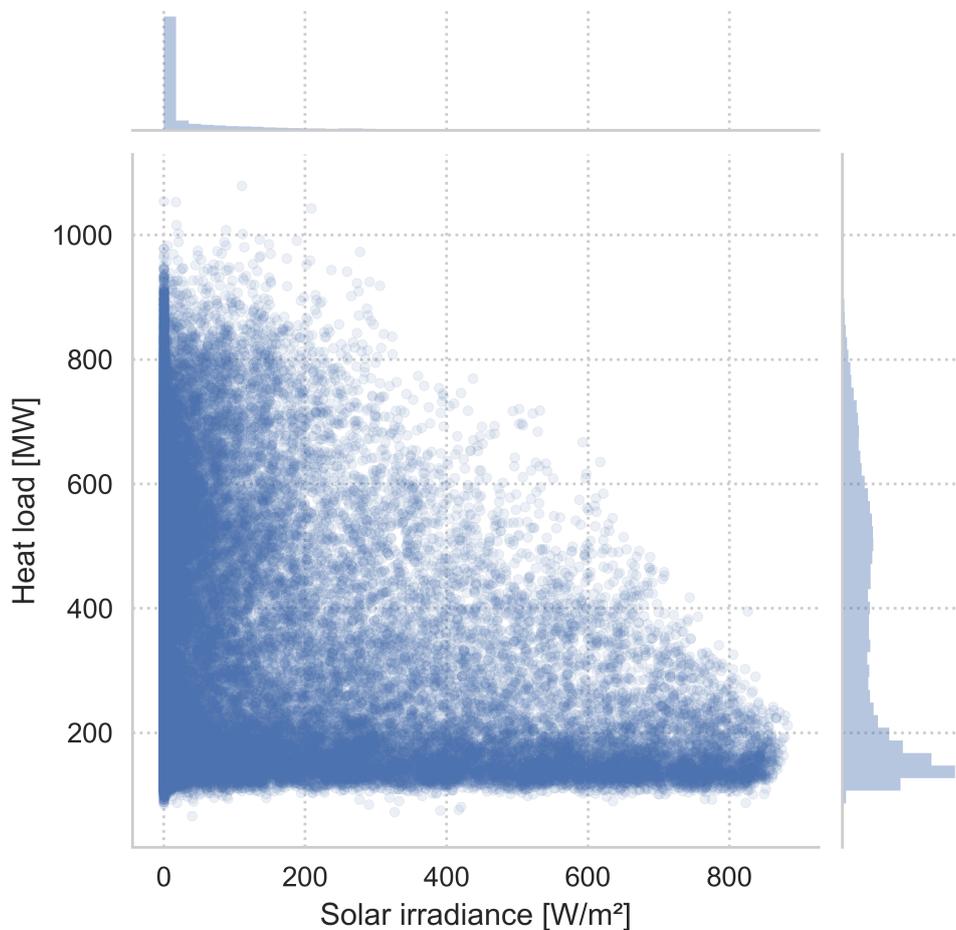
While there is a strong relationship between the heat load and the outside temperature, the point cloud has a substantial vertical width. Therefore, the outside temperature is insufficient in itself to accurately model the heat load on an hourly basis. It is necessary to incorporate more data to accurately forecast hourly heat loads, e.g. information about the autocorrelation of the load or other external variables such as wind speed and solar irradiance.



**Figure 2.5:** Scatter plot of the hourly heat load versus the wind speed over seven years. The histograms in the side panels illustrate the distribution of the variables. The Pearson correlation coefficient is 0.27 between heat load and wind speed.

Figure 2.5 depicts the relationship between the average hourly wind speed and the heat load. The relationship is not nearly as well defined as was the case with the temperature, but there is a positive correlation coefficient of 0.27. Previously, wind chill has been shown to affect the heat load mainly at high loads, and with relatively small effect (Frederiksen and Werner, 2013). In this work, wind components are found to be statistically significant, albeit small in load forecasting models.

Finally, Figure 2.6 shows the relation between the heat load and the solar irradiance. Again, no clear relationship can be observed, although there is a



**Figure 2.6:** Scatter plot of the hourly heat load versus the solar irradiance over seven years. The histograms in the side panels illustrate how the variables are distributed. The Pearson correlation coefficient is  $-0.32$  between heat load and solar irradiance.

negative correlation of  $-0.32$ . Much of the correlation may be explained by the temperature dependence on the load and the low solar gains during winter. Heat load reductions due to solar gains have previously been found to be small (Werner, 1984), but most significant during the spring and fall (Frederiksen and Werner, 2013).

Among district heating operators, however, it is well known that the heat load can shift significantly due to changing cloud cover, and it is worth exploring whether solar irradiance can improve heat load forecasts. Including solar gains in heat load forecasts is challenged by the fact that cloud cover is difficult to forecast accurately, especially on very local scale with high temporal resolution.

In the next chapter, I dive into weather-based uncertainties, their influence on heat load forecasts and how better knowledge of these uncertainties can improve operation of district heating systems.



## Chapter 3

# Weather-based forecast uncertainties

### 3.1 Motivation

Operation and planning of energy systems are highly dependent on the weather. In energy systems with large penetrations of wind and solar generation, the supply side of the system is dominated by the fluctuations in wind speeds and cloud cover. On the demand side, especially the heating load, but also the electricity load on the system is highly temperature dependent. The close connection between weather conditions and the operation of energy systems, is bound to pose challenges for system operators and production planners in the electricity and district heating sectors. In this first study, I explore the potential for optimizing system operations through better knowledge of the weather-based risk profile.

Weather prediction is challenging, because it depends on highly complex atmospheric systems. Fortunately, the tremendous societal values of weather predictions have ensured that significant research effort and funding have been devoted to improving weather forecast throughout the 20th century (Inness and Dorling, 2012). Accurately forecasting the weather using numerical weather prediction (NWP) models requires good knowledge about initial conditions as well as accurate numerical representation of the atmosphere. By running NWP models multiple times with perturbed initial conditions and model physics, meteorologists can create a probabilistic forecast called an ensemble forecast. Ensemble forecasts usually consist of 5-100 point forecasts, and can be transformed to probability densities through statistical post-processing (Gneiting and Raftery, 2005).

Today, energy systems researchers can stand on the shoulders of giants, and use the meteorological estimates of the uncertainties of weather forecasts to explore the impact of imperfect knowledge on energy systems. Electricity load forecasts have been extended using ensemble weather prediction to create probabilistic load forecasts (Taylor and Buizza, 2002, 2003). Ensemble weather predictions have also been used to estimate the uncertainty of wind (Möhrlen and Jørgensen, 2006; Taylor et al., 2009) and solar (Alessandrini et al., 2015) power forecasts. However, ensemble weather predictions have not previously

been applied to district heating load forecasting.

The following article, published in the Elsevier journal *Applied Energy*, demonstrates the value of using ensemble weather predictions for heat load forecasting. This first use-case is focused on the distribution part of a district heating system and how better knowledge of the forecast uncertainties can save energy by reducing the heat losses to the ground.

## 3.2 Methods

The methodology of the study can be divided into two parts. In the first part, I apply ensemble weather predictions to heat load forecasting. In the second part I simulate how the information about the forecast uncertainty can be applied to improve the control of the distribution of heat to consumers.

### Ensemble heat load forecast

The heat load in district heating systems has a strong 24 h autocorrelation. Starting from this point, and correcting for the changes in the weather, I have constructed a simple autoregressive model to forecast the heat load, based on the outside temperature, wind speed and solar irradiance. The model reads

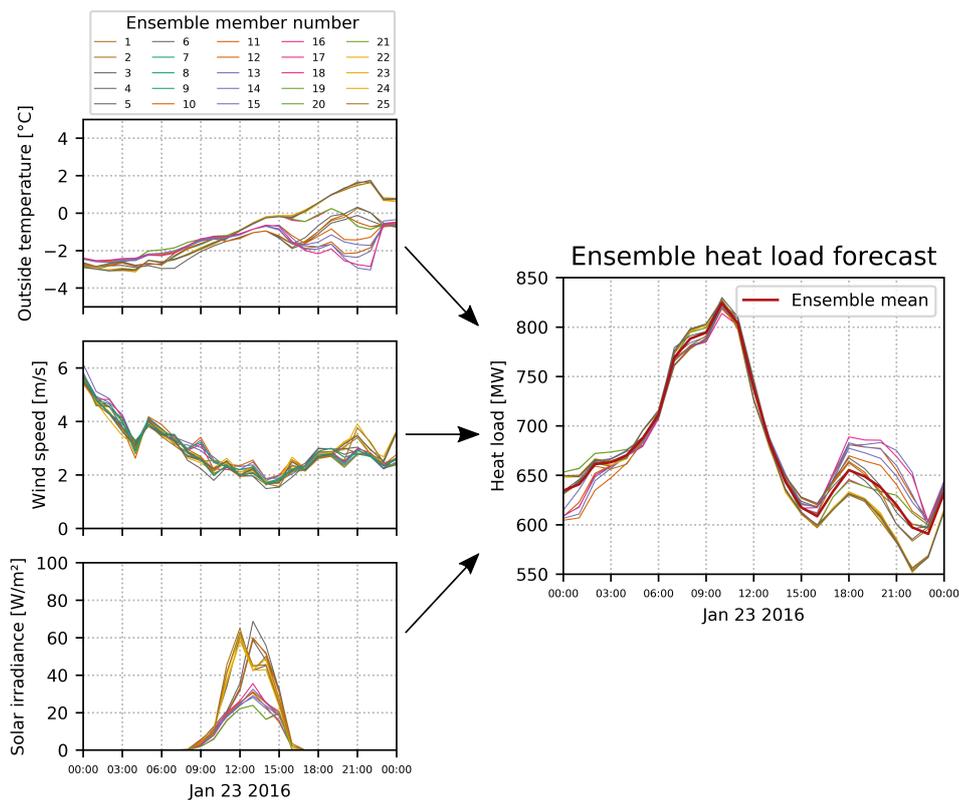
$$\begin{aligned} \hat{P}_{t|t-24} = & a P_{t-24} + b (\hat{T}_{t|t-24}^{\text{out}} - T_{t-24}^{\text{out}}) \\ & + c (\hat{v}_{t|t-24}^{\text{wind}} - v_{t-24}^{\text{wind}}) + d (\hat{I}_{t|t-24}^{\text{sun}} - I_{t-24}^{\text{sun}}), \end{aligned} \quad (3.1)$$

where  $P$  is the heat load and  $t$  denotes the hourly time steps. The weather parameters outside temperature, wind speed and solar irradiance are denoted by  $T^{\text{out}}$ ,  $v^{\text{wind}}$  and  $I^{\text{sun}}$ . Predicted values of load and weather variables are accented by  $\hat{\cdot}$  and  $a$ ,  $b$ ,  $c$  and  $d$  are model parameters. The predicted heat load is dominated by the outside temperature and the 24 h-lagged value of the heat load. The intuitive interpretation of the model is that the heat load is forecasted as the 24 h-lagged heat load corrected for the change in weather conditions over the past 24 h.

In order to create an ensemble of heat load forecasts, the model (3.1) has been fed with an ensemble of weather forecasts. The Danish Meteorological Institute (DMI) has kindly provided me with ensemble forecasts consisting of 25 instances of weather forecast for the city of Aarhus. Feeding these 25 ensemble members through the heat load forecast model resulted in an ensemble of 25 heat load forecasts, as illustrated in Figure 3.1. From the spread of the load forecast ensemble, the time-dependent weather-based forecast uncertainty can be estimated. The higher the spread of the ensemble forecast, the higher the uncertainty.

### Uncertainty-based temperature control

In large district heating systems, the heat is distributed to the consumers through a number of area substations equipped with heat exchangers. The heat



**Figure 3.1:** Example of how ensemble weather predictions are combined to create an ensemble heat load forecast. The forecast for January 23 2016 is shown, and it is calculated at midnight using the model in (3.1).

delivered across a heat exchanger per unit time  $P$  is given as

$$P = c_w \rho_w Q (T_{\text{sup}} - T_{\text{ret}}) , \quad (3.2)$$

where  $T_{\text{sup}}$  and  $T_{\text{ret}}$  are the supply and return temperatures.  $Q$  is the volume flow rate and  $c_w$  and  $\rho_w$  are the specific heat and density of water.

The delivered heat can be increased in two ways. The first is by raising the supply temperature, and thus increasing the gap between supply and return temperatures, since the return temperature is largely outside of the control of the operator. The second way is by increasing the volume flow rate.

Raising the supply temperature increases the amount of heat that is lost to the ground. A rule of thumb among operators says that this heat loss is an economical order of magnitude larger than the extra pumping power required when increasing the volume flow rate instead. Thus, it is worthwhile to keep the supply temperature as low as possible, but this can cause problems in a peak load situation. If the system runs out of pumping capacity, because the supply temperature is too low, severe operational problems follow.

Therefore, it is customary to build a security margin into the supply temperature control. The security margin accounts for the uncertainty in the load

forecast, which means that it can be narrower when the forecast uncertainty is small and wider when the forecast is very certain.

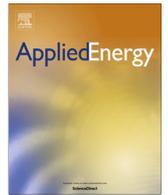
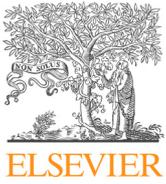
In the paper, I compare the performance of an optimized temperature control scheme in three situations: The current operation of the substations, security margins based on constants forecast uncertainties and security margins based on time-dependent weather-based uncertainties.

### 3.3 Main findings

For the first time, I demonstrated that probabilistic heat load forecasts can be created in a straight-forward way based on ensemble-weather predictions. This technique can provide operators and production planners of district heating systems with valuable knowledge about the time-dependence of the forecast uncertainty. With this in hand, operators know when to play it safe, and when they can push the system.

As a first application of ensemble heat load forecasts, I explored the supply temperature control of area substations. Controlling the supply temperature according to time-dependent weather based forecast uncertainties had a small potential to reduce heat losses and save energy. However, in systems with under-dimensioned, or limited pumping capacity, more significant benefits could be observed.

The low-level operation of district heating systems can benefit, not only from better predictions of load, but from better knowledge about the *uncertainty* of those predictions.



# Using ensemble weather predictions in district heating operation and load forecasting



Magnus Dahl<sup>a,b,\*</sup>, Adam Brun<sup>b</sup>, Gorm B. Andresen<sup>a</sup>

<sup>a</sup> Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark

<sup>b</sup> AffaldVarme Aarhus, Municipality of Aarhus, Bautavej 1, 8210 Aarhus V, Denmark

## HIGHLIGHTS

- Ensemble weather predictions are introduced in district heating operation.
- A heat load forecast model with dynamic weather-based uncertainties is developed.
- Dynamic forecast uncertainties are applied to the operation of area substations.
- The supply temperature can be lowered while retaining security of supply.
- Area substations with smaller pumping capacity benefit most.

## ARTICLE INFO

### Article history:

Received 13 October 2016

Received in revised form 9 February 2017

Accepted 22 February 2017

### Keywords:

District heating

Ensemble weather data

Heat load forecasting

Operation under uncertainties

## ABSTRACT

Ensemble weather predictions are introduced in the operation of district heating systems to create a heat load forecast with dynamic uncertainties. These provide a new and valuable tool for time-dependent risk assessment related to e.g. security of supply and the energy markets. As such, it is useful in both the production planning and the online operation of a modern district heating system, in particular in light of the low-temperature operation, integration of renewable energy and close interaction with the electricity markets. In this paper, a simple autoregressive forecast model with weather prediction input is used to showcase the new concept. On the study period, its performance is comparable to more complex forecast models. The total uncertainty of the heat load forecast is divided into a constant model uncertainty plus a time-dependent weather-based uncertainty. The latter varies by as much as a factor of 18 depending on the ensemble spread. As a consequence, the total forecast uncertainty varies significantly. The forecast model is applied to the operation of three heat exchanger stations. Applying an optimized temperature control can significantly lower supply temperatures compared to current operation. Improving the temperature control with dynamic time-dependent weather-based uncertainties can lower the supply temperature further and reduce heat losses to the ground. The potential benefit of using dynamic uncertainties is larger for systems with relatively smaller pumping capacities.

© 2017 Elsevier Ltd. All rights reserved.

## 1. Introduction

District heating systems exist in most countries in the Northern Hemisphere but are most widespread in the Nordic countries and in the former Soviet Union [1]. In the EU, district heating covers about 13% (2010) of the total domestic heating demand. It has been estimated that this could potentially be increased to 50% by 2050 [2]. Unlike individual house heating, district heating requires

investment in city-wide distribution networks. As a consequence, district heating is not competitive in low density areas, but has a significant potential in many high heat density urban areas, despite reduced heat demand from future building energy retrofit solutions [3].

District heating can limit the use of high exergy fuels such as oil and gas for heat-only applications. This is due to its ability to utilize low-quality energy sources such as municipal waste and excess heat from heavy industrial processes and electricity production. Increasing the use of biomass and solar thermal energy in the district heating sector is useful in the process of decarbonizing the energy sector. Combined use of large-scale heat pumps and electric boilers allows district heating systems to utilize electricity

\* Corresponding author at: Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark.

E-mail address: [magnus.dahl@eng.au.dk](mailto:magnus.dahl@eng.au.dk) (M. Dahl).

## Nomenclature

$\delta\hat{P}$	uncertainty on the heat demand forecast $\hat{P}$ [MW]	$Q^{\text{ref}}$	reference volume flow rate for security of supply [ $\text{m}^3/\text{h}$ ]
$\hat{P}$	forecasted heat load, production or consumption [MW]	$Q_{\text{max}}$	water flow capacity of a heat exchanger station [ $\text{m}^3/\text{h}$ ]
$\mathcal{L}$	likelihood function	$rh$	relative humidity [%]
$\rho_w$	density of water [ $\text{kg}/\text{m}^3$ ]	$t$	subscript denoting hourly time steps
$\sigma$	standard deviation of forecast errors [MW]	$T^{\text{out}}$	outside temperature [ $^{\circ}\text{C}$ ]
$\sigma^{\text{m}}$	standard deviation of model-based errors [MW]	$T_{\text{ret}}$	return temperature [ $^{\circ}\text{C}$ ]
$\sigma_t^w$	standard deviation of the ensemble of heat demand forecasts in time step $t$ [MW]	$T_{\text{sup}}$	supply temperature [ $^{\circ}\text{C}$ ]
$\sigma_t^{\text{tot}}$	combined standard deviation, based on $\sigma_t^w$ and $\sigma^{\text{m}}$ in time step $t$ [MW]	$T_{\text{sup}}^{\text{min}}$	minimum supply temperature as a function of outside temperature [ $^{\circ}\text{C}$ ]
$a, b, c, d$	heat load model parameters	$v^{\text{wind}}$	wind speed [m/s]
$c_w$	specific heat of water [MW h/kg. $^{\circ}\text{C}$ ]	AIC	Akaike Information Criterion
$I^{\text{sun}}$	solar irradiance [ $\text{W}/\text{m}^2$ ]	MAE	mean absolute error [MW]
$k$	number of model parameters, used when evaluating AIC	MAPE	mean absolute percentage error [MW]
$P$	heat load, production or consumption [MW]	RMSE	root mean square error [MW]
$Q$	volume flow rate [ $\text{m}^3/\text{h}$ ]	Sc. $i$	superscript denoting Scenario $i$

generated from wind and solar power for heating in situations with surplus generation [4,5].

Forecasting of both production and demand is becoming increasingly important in the energy sector due to: (i) the growing share of wind and solar energy and (ii) the focus on coupling the electricity, heating and transportation sectors in smart energy networks [6]. In this context, a good forecast can be used to plan the operation of flexible assets to minimize costs and environmental impact. The uncertainty of the forecast can be used to quantify financial risk in the energy market or operational risk related to security of supply [7]. Forecast uncertainty can be estimated using the technique of ensemble forecasting. Ensemble forecasting has previously been used for electricity load forecasting, both in a linear [8] and in a neural networks context [9]. It has also been applied to wind [10,11] and solar power forecasting [12] to estimate the forecast uncertainty. In this paper, the technique of ensemble forecasting is adapted to district heating load forecasting for the first time. We are also the first to demonstrate how dynamic forecast uncertainty could be applied to operate existing heat exchanger stations with increased efficiency while retaining security of supply.

### 1.1. Heat load forecast uncertainty

District heating production planning and operation involve decision making under uncertain conditions. Hence, accurate forecasts of daily variations in the heat load are needed in the district heating sector.

Variations in the heat load are caused by changing weather and consumer behavior [13], and forecasting district heating demand or heat load has been studied in a multitude of papers. A number of commercial tools for this purpose also exist. Within academic studies, machine learning approaches to top-down forecasting of district heating demand have gained popularity in recent years. Multilayered neural networks are used to predict heat load in [14,15], and to predict cooling load in a district heating and cooling system in [16]. In [17], the authors compare a number of different supervised machine learning algorithms and conclude that support vector regression performs best. Other studies such as [18] take a more traditional statistical approach. In [18], physical knowledge is used to limit the model space and statistical analysis is used to refine the model and estimate the parameters. Stochastic time series methods have also been successfully applied to heat load forecasting. This includes general transfer function models [19] and

seasonal autoregressive integrated moving average models (SAR-IMA) [20,21]. The paper [22] demonstrated that decent forecast performance can be achieved by simple autoregressive methods with weather input. In [21], a comparison between a number of linear regression models and a SARIMA model with exogenous input favors a simple linear regression model using weekly heat demand patterns. On the building level, [23] presents a heat demand forecast that is a hybrid model using both physical, autoregressive integrated moving average models (ARIMA) and singular value decomposition methods.

While all of these studies benchmark the performance of their models by some standard measure, not all of them address the forecast uncertainty directly, e.g. through the use of prediction intervals. In [20,23], prediction intervals are provided for the forecast models, but these are estimated from the statistical uncertainty on the model parameters only, and they do not take the unpredictability of the weather input into account. In [21,22], for instance, a perfect weather forecast is assumed when benchmarking the model. However, weather forecasts are never perfect, and a heat load forecast model that depends on a weather forecast is bound to propagate some of the uncertainty in the weather forecast into an uncertainty on the heat load forecast. The authors of [15] estimate weather-based prediction intervals for one of their forecast models. These prediction intervals are constant in time and are estimated by simply adding Gaussian noise with mean 0 and standard deviation 1 to each of the weather variables in a Monte Carlo simulation.

In this paper, we present a heat load forecast model with prediction intervals that vary in time, due to the time variation in the uncertainty of an ensemble weather forecast. An ensemble weather forecast consists of multiple independent forecasts that ideally cover the full range of possible weather conditions in a given period. It can be used to estimate the most likely scenario and quantify the time-dependent uncertainty of the weather variables. Each ensemble member represents an internally consistent weather configuration based on a sophisticated meteorological model which means that cross-correlations between different weather variables are naturally captured.

A heat load forecast with dynamic prediction intervals is a new and valuable tool for the district heating sector. Augmented with cost estimates, it will allow production planners to take calculated risk in unit commitment situations or when trading on the electricity market. From an operational perspective, knowing the dynamic uncertainties of a forecast enables operators to know when to push

the equipment and when to exercise caution in uncertain weather situations.

## 1.2. Application case

The potential savings of lowering the supply temperature are significant as heat losses to the ground are reduced. In a district heating system such as Aarhus' the total annual losses amount to approximately 15–20% of the produced heat, and reducing it is very costly. In smaller systems, the relative heat losses can be even larger. Heat losses from district heating systems to the ground have been studied thoroughly in the literature both from a steady state [24,25] and from a transient perspective [26]. Additional benefits of lowering the supply temperature are widely known, and it is a key element in the future's 4th generation district heating (4GDH) [27]. Other important benefits of lowering supply temperature include: more electricity generated from CHP plants at the same heat demand, higher COP value for large heat pumps and more heat to recover from flue gas condensation [28].

A number of studies focus on dynamic optimization of different district heating systems. In an early study [29], a model intended for online unit commitment and temperature control is described. More sophisticated hydraulic calculations are included in a number of similar, later studies. The most recent include [30] which focuses on optimal pump control and computational speed, [31] where a number of renewable heat sources are included in addition to traditional heat plants and [32] where a circular network topology and a new mass flow control scheme is investigated. In all cases, a central constraint in the optimization is a fixed maximum volume flow. In real systems, however, a safety margin is imposed on the maximum flow by the operator to be able to react to unforeseen increases in consumption. The magnitude of such events depends on the heat load forecast error, part of which originates from uncertainties in the weather forecast. In the second part of this paper, we demonstrate how heat exchanger stations with dynamic security margins corresponding to the dynamic uncertainties of the ensemble based heat load forecast can be operated to lower the supply temperature in peak load situations. This coupling between forecast uncertainty and operational risk assessment in district heating has not previously been studied.

The paper is structured as follows. Section 2 introduces the ensemble weather data and describes and validates a heat load forecast model. In this section we apply ensemble predictions to provide a heat load forecast with dynamic prediction intervals. Section 3 deals with the application case and describes the benefits of operating heat exchanger stations according to ensemble forecasts. Section 4 concludes the paper.

## 2. Heat load forecast with weather ensemble input

The foundation of this analysis is an ensemble of 25 weather forecasts for a geographical point in Aarhus. These weather forecasts are used as input to a simple model for the total production of heat needed in Aarhus. The output is an ensemble of 25 production time series, allowing us to estimate the uncertainty of the heat load forecast. This uncertainty is composed of a weather-based uncertainty and a model uncertainty.

### 2.1. Ensemble weather forecast data

The forecast ensemble consists of 25 ensemble members with hourly values of outside temperature, wind speed, solar irradiance and relative humidity. The forecasts are based on the HIRLAM numerical weather prediction (NWP) model [33] and provided by the Danish Meteorological Institute (DMI). The ensemble members

are generated by perturbing the initial conditions of the NWP model with the purpose of estimating the probability distribution of the forecasted weather [8]. Each forecast has a horizon of 54 h and a new forecast becomes available every 6 h. The spatial resolution of the forecast is about 5.5 km × 5.5 km. In this study, the ensembles for a geographical point in the western part of Aarhus are used (N 56°10'22.6", E 10°8'2.2"). Forecast data ranging from December 16th 2015 to March 1st 2016 has been available for this analysis.

For simplicity, the many overlapping ensemble time series have been reduced to a single ensemble of time series of outside temperature, wind speed and solar irradiance. The value of the most recent forecasts was assigned to each hour. Due to a 5 h delay in the availability of the most recent forecast, the hours 5–11 from each set of ensembles are used to resemble the most recent data that would be available in real time operation. Fig. 1 shows the resulting 25 time series for outside temperature, wind speed and solar irradiance for a period in January and February 2016. Due

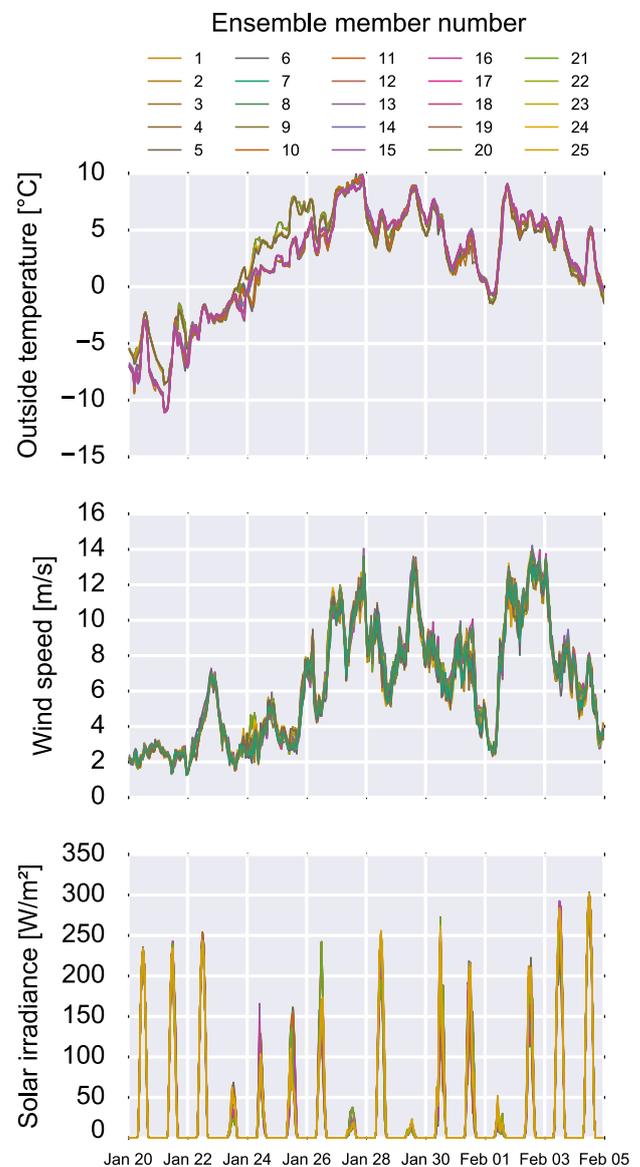


Fig. 1. Time series for the most recent available forecast for 25 ensemble members. Outside temperature, wind speed and solar irradiance are shown with hourly resolution for the period January 20th to February 5th 2016. Relative humidity is not shown, as it proved irrelevant in the heat load modeling.

to the short forecast horizons, many of the ensemble members are fairly close to each other, thus making it difficult to see how many there are. However, it is evident that the ensemble spread is not constant, and this variation is crucial to the subsequent analysis, e.g. notice how the temperature ensemble divides into two clusters between January 24th and 26th.

## 2.2. Heat load forecast model

The heat load of Aarhus is the total heat production from all the heat producing plants in the Aarhus district heating system. We have used measured hourly production data for the period December 16th 2015 to March 1st 2016, except for January 15th to 20th 2016, due to a data outage in the SCADA system. To forecast the heat demand, we used weather forecasts from DMI, available at the time of prediction, as opposed to historical weather data.

For this paper, a simple and intuitive class of production forecast models has been selected to high-light the new concept of ensemble forecasting in district heating applications. The production in a given hour is predicted as the production 24 h before, adjusted by the change in weather conditions. This is a special case of an autoregressive model with exogenous input (ARX). Mathematically it reads:

$$\hat{P}_{t|t-24} = aP_{t-24} + b(\hat{T}_{t|t-24}^{\text{out}} - T_{t-24}^{\text{out}}) + c(\hat{v}_{t|t-24}^{\text{wind}} - v_{t-24}^{\text{wind}}) + d(\hat{I}_{t|t-24}^{\text{sun}} - I_{t-24}^{\text{sun}}). \quad (1)$$

Here  $P$  is the total production of heat per unit time and  $\hat{\cdot}$  denotes the predicted value. The subscripts denote hourly time steps and  $T^{\text{out}}$ ,  $v^{\text{wind}}$  and  $I^{\text{sun}}$  denote outside temperature, wind speed and solar irradiance respectively.  $a$ ,  $b$ ,  $c$  and  $d$  are model parameters. Note that  $a$  has no physical interpretation and depends on the time series only. In order to ease the notation, we suppress the prediction horizon and let  $\hat{P}_t$  denote predicted heat load in time step  $t$  based on the most recent available weather forecasts.

### 2.2.1. Model selection

The ensemble weather forecast included the following weather variables: outside temperature, wind speed, solar irradiance and relative humidity. For model selection and fitting the model the ensemble mean of each of the weather variables was used. This technique often provides a better weather forecast compared to using a single run of a numerical weather model [34]. The appropriate model structure (1) was determined in the following way: we began with the simple assumption that  $\hat{P}_t = aP_{t-24}$ . This assumption was motivated by the fact that the production time series has a clear daily pattern and a strong 24 h autocorrelation. Then we constructed a multitude of models by adding all the possible combinations of 24 h differenced weather terms such as  $b(\hat{T}_t^{\text{out}} - T_{t-24}^{\text{out}})$ , correcting for the weather changes during the past 24 h. This process yielded 16 different model structures. The models were fitted and cross-validated on the period December 17th to February 5th (except January 15th to 20th due to an outage in the SCADA system). In the model selection and benchmarking process we used 10-fold cross-validation. Two criteria were used for model selection. The first was the root mean square error. It is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_t (\hat{P}_t - P_t)^2}, \quad (2)$$

where  $N$  is the number of time steps and  $\hat{P}_t$  and  $P_t$  are the predicted and actual production in time step  $t$ . The root mean square error was calculated for all the validation periods and averaged. Of all the 16 models, the one described in (1) yielded the lowest root

mean square error. The second model selection criterion was the Akaike Information Criterion (AIC), defined as

$$\text{AIC} = 2k - \ln(\mathcal{L}), \quad (3)$$

where  $k$  is the number of free parameters in the model, and  $\mathcal{L}$  is the likelihood function [35]. The AIC is a measure of the information that is lost when representing a signal with a statistical model; therefore the lower the AIC-value is, the better is the model. The AIC rewards goodness of fit, but punishes the number of variables, and is a good criterion for avoiding overfitting.

Models including both temperature  $b(\hat{T}_t^{\text{out}} - T_{t-24}^{\text{out}})$  and wind  $c(\hat{v}_t^{\text{wind}} - v_{t-24}^{\text{wind}})$  yielded reasonable results. Adding either a solar term  $d(\hat{I}_t^{\text{sun}} - I_{t-24}^{\text{sun}})$  or a humidity term  $e(\hat{r}h_t - rh_{t-24})$  improved the model slightly, but the one with the solar term performed best. These two models and the one with all four weather differenced terms had very similar AIC-values. Thus, we settled on the model in (1) with no humidity term as, it had the lowest RMSE and a very low AIC-value in the cross-validation. Another problem with the model including all the weather terms was that the  $p$ -value for the humidity terms was often so high that the null hypothesis that the coefficient was 0 could not be rejected at the 5% confidence level.

The model parameters for (1) estimated from the cross-validation period are shown in Table 1. Fig. 2, shows an example of the prediction of the model along with the historical production for the period January 20th to February 5th. The black curve is the historical production and the red curve is the model. The prediction intervals shown in blue and gray will be discussed in Section 2.3.

### 2.2.2. Model performance

Three standard error measures were used to benchmark the performance. Besides the root mean square error we used the mean absolute error defined as

$$\text{MAE} = \frac{1}{N} \sum_t |\hat{P}_t - P_t|, \quad (4)$$

and the mean absolute percentage error given by

$$\text{MAPE} = \frac{1}{N} \sum_t \left| \frac{\hat{P}_t - P_t}{P_t} \right|. \quad (5)$$

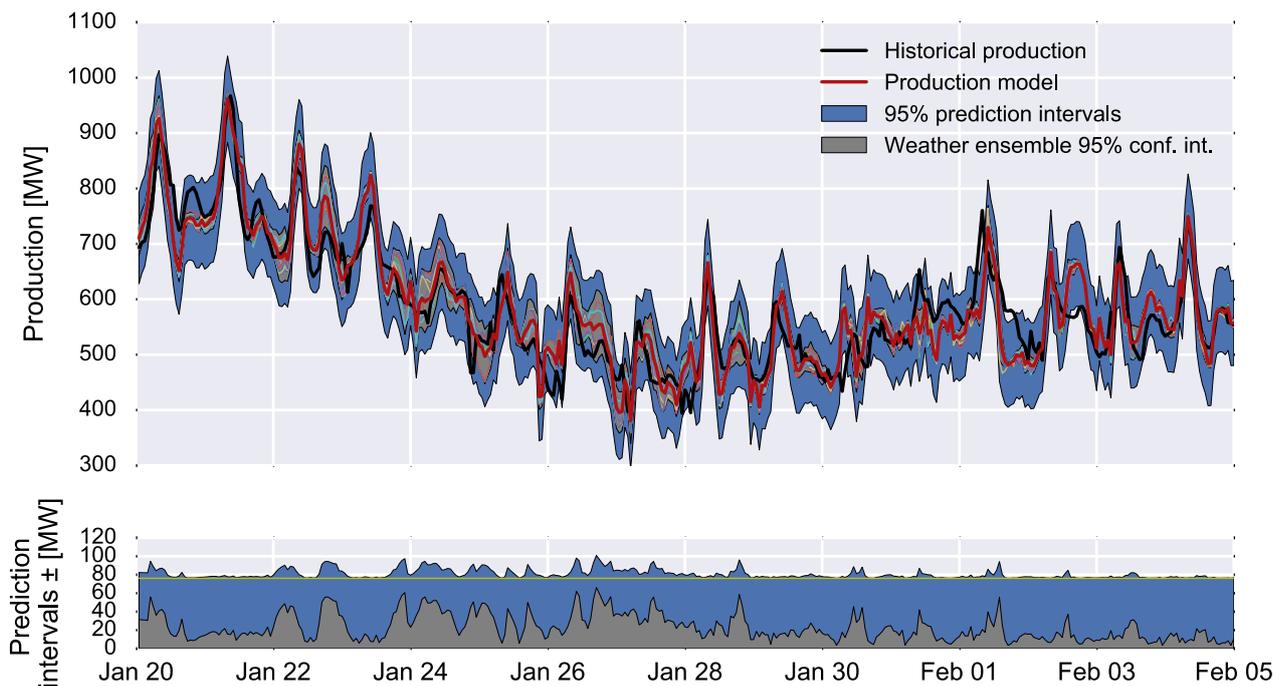
The model's performance was evaluated through 10-fold cross-validation on the cross-validation period. We also evaluated the model's forecast performance in an out-of-sample test using the period from February 5th to March 1st. Note that this data has not been used for fitting or model selection and is completely blind.

The model's performance in the cross-validation is comparable to the performance on the blind test period. Furthermore, its performance is comparable to that of a commercial forecasting system utilized by the district heating company of Aarhus on a forecast horizon of 0–24 h. The performance of our model in the cross-validation and test period is summarized in Table 2. For comparison, we also show the performance of the commercial forecasting

**Table 1**

Parameter estimates from the model (1) fitted to production data from December 17th 2015 to February 5th 2016.

	Estimate	Standard deviation	Unit
$a$	0.999	0.002	–
$b$	–20.0	0.5	MW/°C
$c$	6.1	0.4	MW/m/s
$d$	–0.21	0.03	MW/Wm <sup>2</sup>



**Fig. 2.** The production model on the period from January 20th to February 5th along with the actual production. The width of the gray band shows the weather-based 95% confidence bounds. The combined width of the gray and the blue band indicates the total 95% prediction interval, including both weather-based uncertainties and model uncertainties. On the bottom the prediction interval is shown. The pure model uncertainty  $1.9599\sigma^m$  is drawn as a fine yellow line. The very fine colored curves in the top figure are the production ensemble members used to estimate the weather-based uncertainty. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

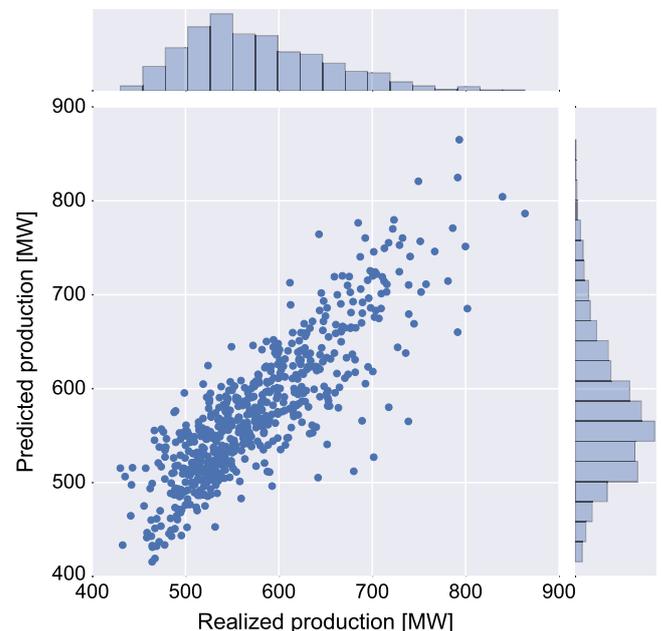
Root mean square error, mean absolute error and mean absolute percentage error in the cross-validation and on the test period for the presented model (1) and the commercial system used in the production planning of the Aarhus district heating system.

	RMSE (MW)	MAE (MW)	MAPE (%)
Cross-validation	41	32	5.7
Blind test period	41	31	5.4
Commercial system, blind test period	41	32	5.4

system on the test period. Fig. 3 is a scatter plot of the predicted versus the realized heat load in the blind test period. It is clear that the model predicts the heat demand reasonably well without over-fitting. From the histograms we also see that the main structure of the distribution of the heat load is captured as well.

For the application of ensemble forecasting to the operation of heat exchanger stations, the model has been used to forecast heat consumption in three different area substations in the Aarhus Area: Holme, Rundhøj and Hørning. The model was fitted, cross-validated and benchmarked as described above. Naturally, this resulted in parameter estimates different from those shown in Table 1 where the model was fitted to the total production, but  $\alpha$  was always very close to 1, indicating a stable model. The forecast performance on each of the different areas is summarized in Table 3. It is clear that the model performance is reasonable and similar across the three different substations.

It is worth noting that the model performs decently, despite its simplicity. This model was chosen for its simplicity and transparency. For online forecasting purposes it is possible to achieve better performance. This can be done by using more complex models such as support vector regression with non-linear kernels and by continuously refitting the model to account for the newest



**Fig. 3.** Scatter plot of the predicted heat load versus the realized heat load on the blind testing period.

available data. However, such an approach would be less transparent and reproducible for the present proof-of-concept application. We emphasize that this model is valid for the winter period only and is intended for demonstration purposes. Due to the limited data period, certain seasonal dynamics in the relationship between the heat demand and the weather may not be captured. We therefore recommend that district heating operators who wish to use ensemble based uncertainties in the operation augment their

**Table 3**

Mean absolute percentage error for the cross-validation and on the test period for the model (1) applied to the heat load of the area substations Holme, Hørning and Rundhøj.

Substation	MAPE		
	Holme (%)	Hørning (%)	Rundhøj (%)
Cross-validation	5.6	4.6	5.3
Blind test period	5.3	6.5	6.2

own heat load prediction model with ensemble weather forecasts. The methodology we apply in the following section can be applied to any heat load prediction model that is based on weather forecasts.

### 2.3. Uncertainties in the heat load model

The production model has uncertainties stemming from two sources: From statistical uncertainty of the model and from uncertainty on the weather input.

Feeding the 25 weather ensemble members through the production model yields an ensemble of possible heat load outcomes. The weather-based uncertainty has been estimated by taking the standard deviation of the production ensemble for each time step and scaling it up to a 95% confidence interval under an assumption of normality. This time-varying confidence interval is shown in Fig. 2 as a gray band around the red model curve. If one looks closely, the production ensemble members can be seen as very fine colored curves.

This means that the symmetric, weather-based confidence bounds are

$$\hat{P}_t \pm 1.9599 \sigma_t^w, \quad (6)$$

where  $\sigma_t^w$  is the standard deviation of the ensemble of production models in time step  $t$  and 1.9599 is the 97.5% quantile of the standard normal distribution  $\mathcal{N}(\mu = 0, \sigma = 1)$ .

The combined time-varying 95% prediction interval, including both weather-based uncertainty and statistical model uncertainty has been estimated under the assumption that the weather-based errors and modeling based errors are independent and normally distributed. This means that the combined standard deviation  $\sigma_t^{\text{tot}}$  in time step  $t$ , can be found as

$$\sigma_t^{\text{tot}} = \sqrt{(\sigma_t^w)^2 + (\sigma^m)^2}, \quad (7)$$

where  $\sigma^m$  is the standard deviation of errors stemming from the statistical model uncertainty. It has been estimated such that the number of time steps in which the forecast error exceeds  $\sigma_t^{\text{tot}}$  matches the expected quantile under assumption of normality.

In the bottom part of Fig. 2, the time-variation of the prediction interval is shown. The weather-based part of the prediction interval (in gray) varies between 3.5 MW and 64.9 MW – a variation of more than a factor of 18. This results in a variation of the combined prediction interval between 74.4 MW in the best case and 98.7 MW in the worst case. The constant model uncertainty, plotted as a fine yellow line, is estimated to 74.3 MW. Note that the quadratic summation of the uncertainties makes the model uncertainty dominate completely when the weather-based uncertainty is small.

Applying the combined prediction interval to the period from January 20th to February 5th, we find that  $3.1 \pm 0.1\%$  of the actual historical load lies above the upper bound and  $3.6 \pm 0.1\%$  lies below. Since the upper bound is the most important, and the upper bound is supposed to be a 97.5% quantile, we conclude that this is a reasonable method for estimating the dynamic prediction intervals on the heat load forecast.

In the following section, the procedure outlined above will be applied to estimate the time-dependent weather-based uncertainties on forecasts of the heat consumption on the three heat exchanger stations Holme, Hørning and Rundhøj.

### 3. Application case: heat exchanger operation

The Aarhus district heating system consists of a transmission system and a distribution system. This is a common division in larger district heating systems and is also found in e.g. Copenhagen and Stockholm. The transmission network consists of larger pipes and operates at a high supply temperature compared to the distribution system. Most production units supply their heat into the transmission network. The distribution system is a number of networks that supply heat all the way to the end consumers. The transmission system and the distribution system are connected in a number of area substations or heat exchanger stations where heat is delivered from the transmission side to the distribution side. This is illustrated in Fig. 4. There are around 50 such area substations in the Aarhus area and we have chosen three for this analysis, specifically Holme, Rundhøj and Hørning in the southern part of Aarhus.

In the following sections, the focus is on the distribution side, immediately after the heat exchanger station.

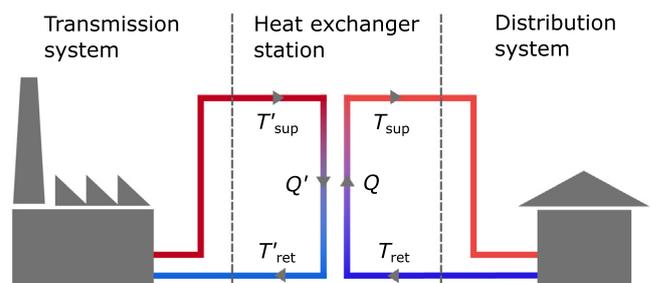
A simple energy consideration shows that the heat delivered per unit time  $P$  across a heat exchanger is

$$P = c_w \rho_w Q [T_{\text{sup}} - T_{\text{ret}}]. \quad (8)$$

Here,  $c_w$  and  $\rho_w$  are the specific heat and density of water at the appropriate temperature,  $Q$  is the water volume flow rate and  $T_{\text{sup}}$  and  $T_{\text{ret}}$  are the supply and return temperatures. This energy balance is true on both sides of the heat exchanger and forms the backbone of the following analysis.

The heat demand on the distribution side is determined by weather and consumer behavior and must be met in all time steps by proper control of the heat exchanger station. The amount of heat delivered to the consumers can be controlled by changing the volume flow rate  $Q$  or the supply temperature  $T_{\text{sup}}$  on the distribution side. The return temperature  $T_{\text{ret}}$  is largely beyond the control of the substation operators and mainly determined by the quality of consumer installations.

From (8) it is clear that the supply temperature can be lowered at the expense of increasing the water flow rate which increases the power consumptions of the pumps. However, this extra cost is small compared to the savings that can be achieved from lowering the supply temperature and will not be considered further.



**Fig. 4.** The transmission system is connected with the distribution system by a number of area substations or heat exchanger stations. The heat consumption from the heat exchanger station is given by (8) in terms of water flow rate  $Q$  and supply and return temperatures  $T_{\text{sup}}$  and  $T_{\text{ret}}$  on the distribution side. The house represents the consumers as well as additional substations on the consumers' premises.

### 3.1. Control scenarios

In the present analysis, we outline and compare three different scenarios for the control and operation of heat exchanger stations. Scenario 1 is the actual historical operation of the stations. Scenario 2 is a control scheme that minimizes the supply temperature while ensuring security of supply. Scenario 3 is the same as Scenario 2 but improved with information about the dynamic weather-based uncertainties of the demand forecast.

Scenario 1 is a reference scenario. It represents the actual historical operation of the heat exchanger stations in the period from December 17th 2015 to March 1st 2016. The three stations are operated differently. Holme and Rundhøj seem to follow a similar scheme in which a supply temperature is set manually to a constant set point and the set point is changed rarely. Hørning is operated by an online installation of the software product Termis<sup>1</sup> and exercises a more aggressive temperature control.

Scenario 2 is an optimized temperature control scenario. It is inspired by the software system PRESS<sup>2</sup> that controls a number district heating systems in Denmark.

The idea is to deliver the demanded amount of heat while minimizing the supply temperature. This must be done under two constraints: (1) The supply temperature must be above a certain level determined by the outside temperature, and (2) security of supply must be ensured.

The minimum allowable supply temperature at a certain outdoor temperature is ultimately an agreement between the consumers and the district heating provider. We determine this function  $T_{\text{sup}}^{\text{min}}(T_{\text{out}})$  as in [19,36]. This is illustrated in Fig. 5, where the supply temperature during each hour over a year is plotted against the low-pass filtered outdoor temperature as in [36].  $T_{\text{sup}}^{\text{min}}(T_{\text{out}})$  (shown in orange) is then chosen such that there is a small (0.5%) chance of falling below this line. Fig. 5 shows data from the Holme station, but similar profiles have been estimated for Hørning and Rundhøj.

Security of supply is ensured by implementing a supply temperature control of the form:

$$T_{\text{sup},t+1}^{\text{sup}} = \max \left\{ T_{\text{sup}}^{\text{min}}(\hat{T}_{t+1}^{\text{out}}), \hat{T}_{t+1}^{\text{ret}} + \frac{\hat{P}_{t+1}}{c_w \rho_w Q_{t+1}^{\text{ref}}} \right\}. \quad (9)$$

Here,  $\hat{P}_{t+1}$  denotes the forecasted heat demand for time step  $t+1$  and  $\hat{\cdot}$  denotes forecasted values. The return temperature is very stable, therefore a reasonable one-step forecast is  $\hat{T}_{t+1}^{\text{ret}} = T_t^{\text{ret}}$ .  $Q_{t+1}^{\text{ref}}$  is a reference flow rate, determined such that the risk of the water flow exceeding the maximum capacity of the pump  $Q_{\text{max}}$ <sup>3</sup> is very small. When the forecast undershoots, the flow rate becomes higher than expected, and it is important to keep the flow rate below  $Q_{\text{max}}$ . If we let  $\delta\hat{P}$  be the uncertainty on the heat demand forecast at the  $3\sigma$  level, there is only a risk of 0.135% of the realized heat demand being larger than  $\hat{P} + \delta\hat{P}$ . This is assuming normality of the forecast errors. We calculate  $Q_{\text{ref}}$  such that the risk of  $Q$  exceeding  $Q_{\text{max}}$  is at this level or smaller. Mathematically, this can be achieved by solving the following minimization:

$$Q_{t+1}^{\text{ref}} = \underset{Q_{\text{ref}}}{\text{argmin}} \left| Q_{\text{max}} - \frac{\hat{P}_{t+1} + \delta\hat{P}}{c_w \rho_w [T_{t+1}^{\text{sup}}(Q_{\text{ref}}) - T_t^{\text{ret}}]} \right|. \quad (10)$$

<sup>1</sup> <http://www.schneider-electric.com/en/product-range/61418-termis-software/>.

<sup>2</sup> <http://www.enfor.dk/products/press.aspx>.

<sup>3</sup> In many cases the maximum capacity is given in terms of a maximum water pressure at a certain point in the system. However, this maximum pressure can be mapped uniquely to a maximum water flow  $Q_{\text{max}}$  with knowledge of the pump and system characteristics.

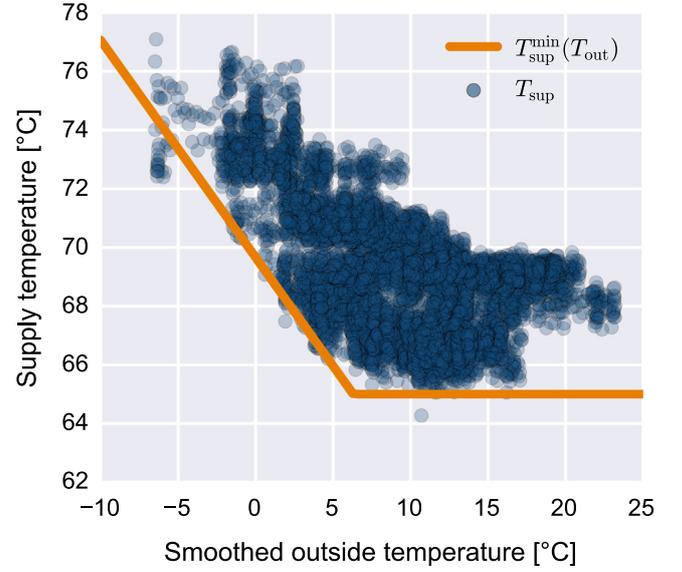


Fig. 5. Supply temperature versus low-pass filtered outside temperature for the Holme station. Each dot represents an hour and one year of data is shown, from March 1st 2015 to March 1st 2016. The minimum supply temperature, as a function of outside temperature is shown in orange.

Here,  $T_{t+1}^{\text{sup}}(Q_{\text{ref}})$  is the supply temperature from (9) with  $Q_{t+1}^{\text{ref}} = Q_{\text{ref}}$ .

This is the optimized temperature control scenario (Scenario 2), where the supply temperature is controlled as in (9) with a reference water flow from (10) and a constant forecast uncertainty of

$$\delta\hat{P} = 3\sigma. \quad (11)$$

$\sigma$  is the standard deviation estimated from the forecast errors.

Scenario 3 is almost identical to Scenario 2, but with the important difference that a time-varying forecast uncertainty is used. In Scenario 3 we set

$$\delta\hat{P}_{t+1} = 3\sigma_{t+1}^{\text{tot}}, \quad (12)$$

where  $\sigma_{t+1}^{\text{tot}}$  is the dynamic weather-based forecast uncertainty as estimated in (7).

### 3.2. Results

In the following sections, we present the potential benefits of using dynamic uncertainties in the operation of heat exchanger stations. The simulation period runs from December 17th 2015 to March 1st 2016 and includes the coldest period in the Danish winter.

#### 3.2.1. Benefit of switch: Scenario 1 → Scenario 2

First, we investigate the potential benefit of switching from today's operation to a scheme with optimized temperature control. Scenario 1 represents actual operational data from the three area substations Holme, Hørning and Rundhøj. Scenario 2 is based on a simulation where the supply temperature is controlled as outlined above while the heat demand of the consumers is the same as in Scenario 1. Discrepancies between the forecasted heat demand and the realized heat demand are covered by adjusting the water flow rate accordingly.

The simulation period covers 1680 h. We use the integrated reduction in supply temperature over the simulation period to evaluate the benefit of switching between scenarios:

$$\int (T_{\text{sup}}^{\text{Sc.1}} - T_{\text{sup}}^{\text{Sc.2}}) dt. \quad (13)$$

We define the sign convention such that a *positive* temperature reduction means that the supply temperature can be *lowered* when switching scenario.

Thus, an integrated temperature reduction of 1680 °C h corresponds to an average supply temperature reduction of 1 °C. In the middle column of Table 4, the benefit of switching from Scenario 1 to Scenario 2 is summarized. Especially the Holme and Rundhøj stations could potentially benefit from an optimized temperature control, with more than 4 °C saved on the average supply temperature. This corresponds to reductions of the heat losses to the ground of 6–7%. For the Hørning station, the benefit of switching to the optimized temperature control of Scenario 2 is smaller, but still significant. This can be explained by the fact that Hørning is already operated by an online Termis installation. Thus, it is expected that the operation of the Hørning station is already somewhat optimized.

Summing up, supply temperatures can be reduced significantly by switching from the current operation to an optimized temperature control.

### 3.2.2. Benefit of switch: Scenario 2 → Scenario 3

We now proceed to investigate the benefit of improving the optimized temperature control with information about the weather-based uncertainties on the demand forecast.

Fig. 6 shows the operational time series of the Holme station during the second half of the simulation period. The top panel shows the heat consumption in black and the forecast of the heat consumption in red. It is clear that the forecast captures the time-like structure of the demand well, and its performance is decent, as summarized in Table 3. The middle panel shows the supply temperature determined by the temperature control (9) in Scenario 2 and Scenario 3. In the bottom panel, the resulting volume flow rate is illustrated. It is obvious that most of the quick fluctuations in demand are covered by similar changes in the water flow. This corresponds well to the fact that the water flow can be changed quicker than the supply temperature. Additionally, very aggressive temperature control can cause mechanical strain on the distribution system.

It is evident that the supply temperature and hence the volume flow rate in many hours are identical in the two scenarios. Utilizing the dynamic weather-based uncertainties, Scenario 3 results in a lower supply temperature in certain peak load hours only. This is more clearly illustrated in Fig. 7. In Fig. 7, the supply temperature reduction  $T_{\text{sup}}^{\text{Sc.2}} - T_{\text{sup}}^{\text{Sc.3}}$ , that is achieved when switching from Scenario 2 to Scenario 3 is shown for the second half of the simulation period. The three area substations Holme, Hørning and Rundhøj are shown in a panel each. In Holme and Rundhøj, the supply temperature can be reduced by approximately 0.5 °C in a number of peak load hours. In Hørning, the potential benefit of using the dynamic weather-based uncertainties of Scenario 3 is more substantial. The supply temperature can be reduced by about 1.2 °C during most hours in the simulation periods and up to 1.5 °C in peak load hours.

The integrated supply temperature reduction in the three areas stations is shown in the right column of Table 4. In terms of supply

temperature reduction, the benefit that can be achieved by switching from Scenario 2 to Scenario 3 is small compared to what can be achieved by the switch from Scenario 1 to Scenario 2. Also, the potential benefit of improving the control with weather-based uncertainties varies greatly from station to station, depending on consumption patterns and the capacity of the pump. However, since Scenario 3 is an improvement of the optimized control in Scenario 2, the benefit of switching to Scenario 3 should be viewed as an extra gain on top of what can be achieved by implementing Scenario 2 alone.

In conclusion, supply temperatures can be reduced additionally by improving an optimized temperature control with dynamic weather-based forecast uncertainties.

### 3.2.3. Sensitivity to pump capacity

The benefit of using dynamic (Scenario 3) versus static (Scenario 2) uncertainties can be divided into two aspects: (1) The number of events in which the two scenarios result in different supply temperatures and (2) the quantity of the supply temperature reduction in individual hours.

The difference between Scenario 2 and Scenario 3 lies in constant versus time-varying forecast uncertainties. Therefore, the only time steps where the two control scenarios result in different supply temperatures are those where the supply temperature depends on the forecast uncertainty *and* the uncertainty is different between the two scenarios. It can be derived from (9) and (10) that the time steps where the supply temperature depends on the forecast uncertainty are exactly those for which the following condition is fulfilled:

$$\hat{P} + \delta\hat{P} > \rho_w c_w Q_{\text{max}} \left[ T_{\text{sup}}^{\text{min}}(\hat{T}^{\text{out}}) - \hat{T}_{\text{ret}} \right]. \quad (14)$$

This simple inequality can provide some insight into which heat exchanger systems could potentially benefit from operating with dynamic weather-based uncertainties as in Scenario 3. If the uncertainty margin  $\delta\hat{P}$  is very large, i.e. if the operator want to use a very high level of security, it is clear that the condition is fulfilled more often. For system operators who want an extremely small risk of the water flow rate exceeding  $Q_{\text{max}}$  this means that there is a larger potential benefit in using dynamic weather-based uncertainties in the operation.

Another crucial variable for the potential benefit of Scenario 3 versus Scenario 2 is the system pump capacity  $Q_{\text{max}}$ . The smaller the system capacity is, the greater the chance that (14) is fulfilled. This results in more hours in which the forecast uncertainty impacts the supply temperature. The magnitude of  $Q_{\text{max}}$  also impacts the other aspect of the size of the benefit: the quantity of the temperature reduction. Differentiating the supply temperature (9) with respect to the forecast uncertainty in the cases where (14) is true yields:

$$\frac{\partial T_{\text{sup}}}{\partial \delta\hat{P}} = \frac{1}{c_w \rho_w Q_{\text{max}}}. \quad (15)$$

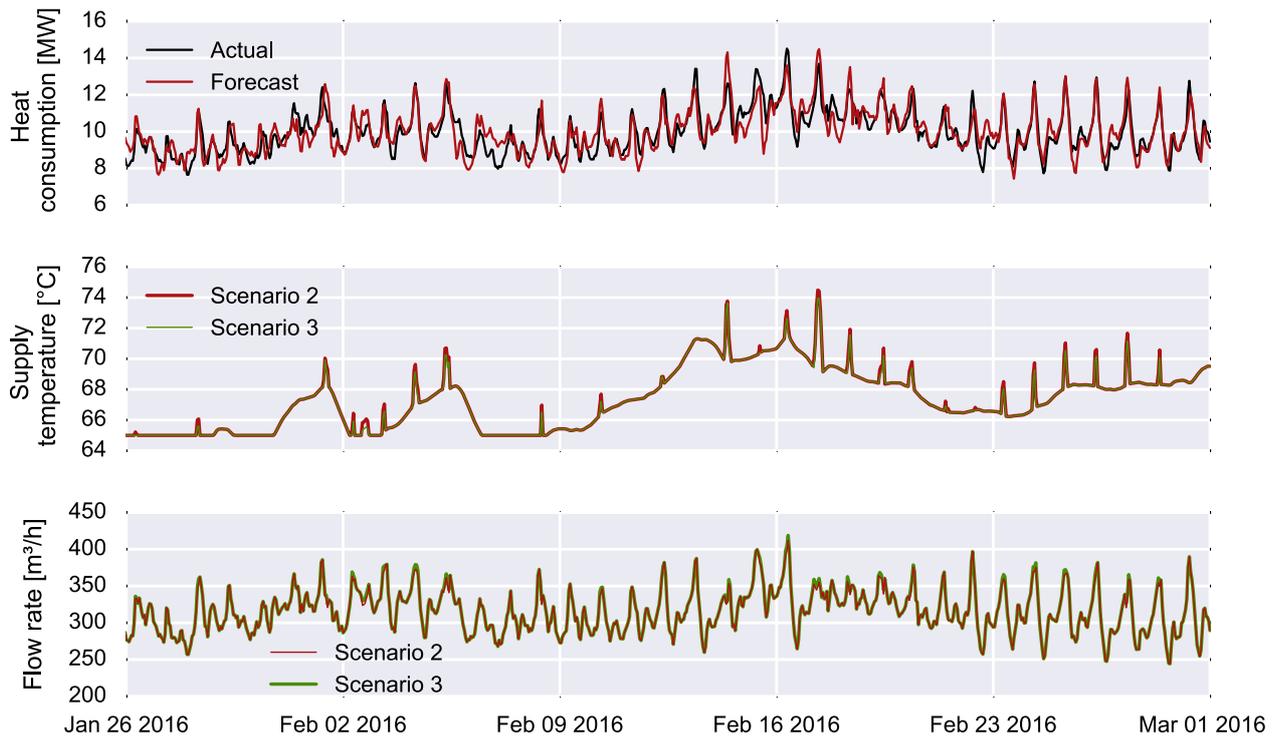
Thus, the supply temperature is a linear function of the magnitude of the uncertainty margin  $\delta\hat{P}$  and the slope is inversely proportional to  $Q_{\text{max}}$ . Therefore, a system with a smaller pump capacity will see larger absolute temperature differences when switching between constant and dynamic uncertainty margins.

To investigate the magnitude of this effect, a simulation has been performed in which the actual maximum capacities for the three heat exchanger stations were scaled down by factors between 1 and 0.5. Fig. 8 depicts the results of this simulation. The number of degree hours of temperature reduction going from Scenario 2 to Scenario 3 is shown as a function of the reduction of the pump capacity. The simulation verifies that the benefit of

**Table 4**

Benefit of switching between the 3 scenarios shown in terms of integrated supply temperature reduction (13) – the number of degree hours saved on the supply temperature. The simulation period spans 1680 h.

	$T_{\text{sup}}$ reduction [°C h]	
	Sc.1 → Sc.2	Sc.2 → Sc.3
Holme	7699	73
Hørning	1606	1333
Rundhøj	7295	27



**Fig. 6.** Scenario 2 and Scenario 3 operational time series for the Holme station in the second half of the simulation period. On the top: Realized heat consumption  $P$  from the station and the forecasted heat consumption  $\hat{P}$  used in the simulation. In the middle: Supply temperature  $T_{sup}$  from the heat exchanger in Scenario 2 and 3. On the bottom: Water flow rate  $Q$  across the heat exchanger in the two scenarios.



**Fig. 7.** Supply temperature reduction when switching from Scenario 2 to Scenario 3:  $T_{sup}^{Sc2} - T_{sup}^{Sc3}$ . This is shown for the three area substations Holme, Hørning and Rundhøj for the second half of the simulation period. For Hørning, the supply temperature in Scenario 3 is higher in a few time steps. This results in negative spikes in mid February reaching values down to  $-0.5$  °C to  $-1.6$  °C.

switching to Scenario 3 is highly dependent on the maximum pump capacity of the stations. Stations with lower capacity will potentially benefit much more from the application of dynamic weather-based uncertainties. Hørning stands out from the two other stations again. This is most likely due to the fact that already

today Hørning is much more pressed for pumping capacity, compared to the other two stations. When dimensioning an area substation, it is important to choose a pump with the correct capacity. If the pump is too large, it will be an unnecessarily large capital investment, and it will operate at a suboptimal efficiency

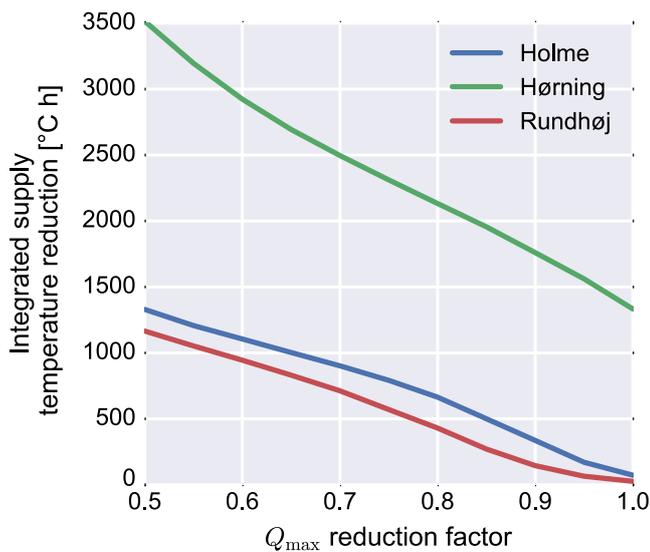


Fig. 8. Benefit of switching from Scenario 2 to Scenario 3 in number of degree hours temperature reduction. This is shown as a function of maximum pumping capacity, going from 50% to 100% of the current value for the three area substations.

for a majority of the time. If the pump is too small, the supply temperature must generally be higher. Therefore, the operational cost of the under-dimensioned systems in the left part of Fig. 8 will generally be higher in terms of heat losses. Thus, switching control scenario cannot be a substitute for correct dimensioning. However, existing under-dimensioned systems can benefit from utilizing dynamic weather-based uncertainties in the operation, as opposed to operation with constant security margins. Using weather-based uncertainties in the operation can limit the added cost of under-dimensioning.

Summing up, the benefit of improving an optimized temperature control with information about dynamic weather-based uncertainties is sensitive to the capacity of the pumping system of the area substation. The smaller the capacity is, the larger the potential benefit is.

#### 4. Conclusion

This paper has introduced the use of ensemble weather predictions in district heating operation by creating a heat load forecast model with time-varying weather-based uncertainties. It has further been demonstrated that information about the time-varying uncertainties can be used to improve an optimized control scheme for heat exchanger stations and lower supply temperatures in the distribution grid. Lowering the supply temperature reduces the amount of heat that is lost from the piping system to the ground.

Based on data from three area substations, it has been found that significant reductions in supply temperature can be achieved by optimizing the current operation of the system with an optimized temperature control. Further but smaller reductions can be achieved by improving the optimized temperature control with dynamic weather-based uncertainties. The potential benefit of using dynamic uncertainties is highly dependent on the pumping capacity of the systems, and systems with smaller capacity can benefit more.

The proposed application case is based on the commercial PRESS software system that is already implemented in a number of district heating systems in Denmark. Improving such an optimal control system or a similar one with dynamic weather-based uncertainties is just a matter of replacing a constant uncertainty

bound  $\delta\hat{P}$  with a dynamic one  $\delta\hat{P}_t$ . The dynamic uncertainty is easily obtained from a heat load forecast when ensemble weather forecasts are provided from the local weather service.

In conclusion, it has been shown how ensemble weather predictions can be used to improve supply temperature control in district heating area substations.

#### Acknowledgements

This study is part of the READY project (Resource Efficient cities implementing ADvanced smart CitY solutions) which is partly financed by the EU's Research and Innovation funding program FP7 ([https://ec.europa.eu/research/fp7/index\\_en.cfm](https://ec.europa.eu/research/fp7/index_en.cfm)). We would like to thank the Danish Meteorological Institute for providing the ensemble weather forecast data and Jens Rishøj Larsen from AffaldVarme Aarhus for fruitful discussions on the operation of heat exchangers.

#### References

- [1] Lundström Lukas, Wallin Fredrik. Heat demand profiles of energy conservation measures in buildings and their impact on a district heating system. *Appl Energy* 2016;161:290–9.
- [2] Connolly David, Lund Henrik, Mathiesen Brian Vad, Werner Sven, Möller Bernd, Persson Urban, Boermans Thomas, Trier Daniel, Østergaard Poul Alberg, Nielsen Steffen. Heat roadmap europe: combining district heating with heat savings to decarbonise the EU energy system. *Energy Policy* 2014;65:475–89.
- [3] Persson Urban, Werner Sven. Heat distribution and the future competitiveness of district heating. *Appl Energy* 2011;88(3):568–76.
- [4] Lund Henrik, Münster Ebbe. Integrated energy systems and local energy markets. *Energy Policy* 2006;34(10):1152–60.
- [5] Meibom Peter, Kiviluoma Juha, Barth Rüdiger, Brand Heike, Weber Christoph, Larsen Helge V. Value of electric heat boilers and heat pumps for wind power integration. *Wind Energy* 2007;10(4):321–37.
- [6] Mathiesen Brian Vad, Lund Henrik, Connolly David, Wenzel Henrik, Østergaard Poul Alberg, Möller Bernd, Nielsen Steffen, Ridjan Iva, Karnøe Peter, Sperling Karl. Smart energy systems for coherent 100% renewable energy and transport solutions. *Appl Energy* 2015;145:139–54.
- [7] Makarov Yuri V, Etingov Pavel V, Ma Jian, Huang Zhenyu, Subbarao Kamesh, et al. Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures. *IEEE Trans Sustain Energy* 2011;2(4):433–42.
- [8] Taylor James W, Buizza Roberto. Using weather ensemble predictions in electricity demand forecasting. *Int J Forecast* 2003;19(1):57–70.
- [9] Taylor James W, Buizza Roberto. Neural network load forecasting with weather ensemble predictions. *IEEE Trans Power Syst* 2002;17(3):626–32.
- [10] Möhrlen Corinna, Jørgensen Jess U. Forecasting wind power in high wind penetration markets using multi-scheme ensemble prediction methods. In: *Proceedings of DEWEK 2006*.
- [11] Taylor James W, McSharry Patrick E, Buizza Roberto. Wind power density forecasting using ensemble predictions and time series models. *IEEE Trans Energy Convers* 2009;24(3):775–82.
- [12] Alessandrini S, Delle Monache L, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. *Appl Energy* 2015;157:95–110.
- [13] Gadd Henrik, Werner Sven. Daily heat load variations in Swedish district heating systems. *Appl Energy* 2013;106:47–55.
- [14] Kato Kosuke, Sakawa Masatoshi, Ishimaru Keiichi, Ushiro Satoshi, Shibano Toshihiro. Heat load prediction through recurrent neural network in district heating and cooling systems. In: 2008 IEEE international conference on systems, man and cybernetics. IEEE; 2008. p. 1401–6.
- [15] Kusiak Andrew, Li Mingyang, Zhang Zijun. A data-driven approach for steam load prediction in buildings. *Appl Energy* 2010;87(3):925–33.
- [16] Sakawa M, Ushiro S, Kato K, Ohtsuka K. Cooling load prediction in a district heating and cooling system through simplified robust filter and multi-layered neural network. *IEEE SMC'99 conference proceedings. 1999 IEEE international conference on systems, man, and cybernetics, 1999, vol. 3. IEEE; 1999. p. 995–1000*.
- [17] Idowu Samuel, Saguna Saguna, Åhlund Christer, Schelén Olov. Forecasting heat load for smart district heating systems: a machine learning approach. In: 2014 IEEE international conference on smart grid communications (SmartGridComm).
- [18] Nielsen Henrik Aalborg, Madsen Henrik. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy Build* 2006;38(1):63–71.
- [19] Madsen Henrik, Sejling Ken, Søgaard Henning T, Palsson Olafur P. On flow and supply temperature control in district heating systems. *Heat Recovery Syst CHP* 1994;14(6):613–20.
- [20] Grosswindhager S, Voigt A, Kozek M. Online short-term forecast of system heat load in district heating networks. In: *Proceedings of the 31st international*

- symposium on forecasting, Prague. International Institute of Forecasters.
- [21] Fang Tingting, Lahdelma Risto. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl Energy* 2016;179:544–52.
- [22] Dotzauer Erik. Simple model for prediction of loads in district-heating systems. *Appl Energy* 2002;73(3–4):277–84.
- [23] Lü Xiaoshu, Lu Tao, Kibert Charles J, Viljanen Martti. Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach. *Appl Energy* 2015;144:261–75.
- [24] Rosa Alessandro Dalla, Li Hongwei, Svendsen Svend. Method for optimal design of pipes for low-energy district heating, with focus on heat losses. *Energy* 2011;36(5):2407–18.
- [25] Li Hongwei, Svendsen Svend. Energy and exergy analysis of low temperature district heating network. *Energy* 2012;45(1):237–46.
- [26] Bøhm Benny. On transient heat losses from buried district heating pipes. *Int J Energy Res* 2000;24(15):1311–34.
- [27] Lund Henrik, Werner Sven, Wiltshire Robin, Svendsen Svend, Thorsen Jan Eric, Hvelplund Frede, et al. 4th generation district heating (4GDH). *Energy* 2014;68:1–11.
- [28] Frederiksen Svend, Werner Sven. District heating and cooling. Studentlitteratur; 2013.
- [29] Benonysson Atli, Bøhm Benny, Ravn Hans F. Operational optimization in a district heating system. *Energy Convers Manage* 1995;36(5):297–314.
- [30] Guelpa Elisa, Toro Claudia, Sciacovelli Adriano, Melli Roberto, Sciubba Enrico, Verda Vittorio. Optimal operation of large district heating networks through fast fluid-dynamic simulation. *Energy* 2016;102:586–95.
- [31] Jiang XS, Jing ZX, Li YZ, Wu QH, Tang WH. Modelling and operation optimization of an integrated energy based direct district water-heating system. *Energy* 2014;64:375–88.
- [32] Laajalehto Tatu, Kuosa Maunu, Mäkilä Tapio, Lampinen Markku, Lahdelma Risto. Energy efficiency improvements utilising mass flow control and a ring topology in a district heating network. *Appl Therm Eng* 2014;69(1):86–95.
- [33] Uden Per, Rontu Laura, Järvinen Heikki, Lynch Peter, Calvo Javier, Cats Gerard, et al. Hirlam-5 scientific documentation. Technical report, Swedish Meteorological and Hydrological Institute; 2002.
- [34] Persson Anders. User guide to ECMWF forecast products. ECMWF; 2015.
- [35] Akaike H. Information theory and an extension of the maximum likelihood principle. In: *Second international symposium on information theory*. p. 267–81.
- [36] Nielsen Torben Skov, Madsen Henrik. Control of supply temperature in district heating systems. In: *Proceedings of the 8th international symposium on district heating and cooling*.



# Chapter 4

## The human factor

### 4.1 Motivation

As engineers and energy system researchers, it is easy to neglect or to forget the factors we cannot control. The previous chapter dealt primarily with how uncertainty about the weather propagates into uncertainty about the load in a district heating system. The weather is one factor that is beyond the control of system operators, human behavior is another such factor.

The load in both electricity and district heating systems has a strong component stemming from the behavior of the consumers. Among district heating operators and production planners, it is well known, that the load on special days such as Christmas or New Year's Eve is difficult to forecast, because people behave differently on these days. On New Year's Eve in Denmark, it is common to go in and out of the house many times during the evening to look at the fireworks. This makes the heat load unusually high, and the consumption pattern notably different from most other days. Human behavior has a significant impact on the operation of energy systems, but this effect has not been covered extensively in energy systems research literature.

This chapter is built around an article published in the special issue of the MDPI<sup>1</sup> journal *Energies: Short-Term Load Forecasting by Artificial Intelligent Technologies*. An earlier version of the paper has appeared in the proceedings for the 12th SDEWES<sup>2</sup> conference.

The article explores the value of including local holiday data in heat load forecasts, taking a first step towards incorporating human behavior into energy systems research.

### 4.2 Methods

In the previous chapter, methods from meteorology were used to quantify the weather-based uncertainty of heat load forecasts. In this chapter, I use methods from machine learning to capture the human component of the heat load. In addition, the present study is based on a much larger data foundation, where

---

<sup>1</sup>Multidisciplinary Digital Publishing Institute.

<sup>2</sup>Sustainable Development of Energy Water and Environment Systems. The 12th conference was held in Dubrovnik.

six years of hourly data was used for training and model selection and one additional year was used for a testing and benchmarking. Finally, this study tests the forecast performance on the time horizons that are necessary for trading decisions on day-ahead electricity markets. This property is highly valuable in modern energy systems where cogeneration and power-to-heat technologies couple the heating and electricity sectors.

### Data scenarios

I grouped the input data, used for load prediction into scenarios. In the first scenario, relevant lags of the heat load as well as the weather variables outside temperature, wind speed and solar irradiance were used for predicting the heat load. In a second scenario, I augmented the lagged heat load and weather variables with data about the hour of day, day of week, weekend and month of the year. Comparison between these two scenarios shows the value of including generic calendar data in the heat load forecasts. In the third scenario, I added data about local national holidays, observances and school holidays. A performance increase, due to the addition of local holiday data, can be attributed to the forecast model better capturing complex consumer behavior around holidays and special occasions.

### Machine learning models

In the paper, I test the forecasting performance of three different machine learning models: ordinary least squares regression (OLS), multilayer perceptron (MLP) and support vector regression (SVR). All three models are well-established in the field of machine learning and have been proven efficient for load forecasting in energy systems. The OLS model is linear, cheap in computation time and a good starting point and reference when comparing with more complex model types. MLP is a simple type of feedforward neural network, and can be extremely powerful at capturing complex input-output relationships. The MLP model, however, requires careful tuning of a number of hyperparameters, such as the number of hidden layers, number of neurons and activation function. The SVR model is a generalization of support vector classification to regression-type problems, and it is capable of modeling nonlinear relationships, through the use of a kernel function. The SVR model also has hyperparameters to tune, although fewer than the MLP model. (Bishop, 2006)

The heat load in district heating systems in temperate climates has a nonlinear dependence on the outside temperature (Frederiksen and Werner, 2013), it is therefore likely that the nonlinear models will show the best performance if they are tuned correctly.

The hyperparameters of the MLP and SVR models have been tuned using  $k$ -fold cross-validation on six years of data. The folds were chosen so each fold contained an entire year, and thus represented the entire annual variation in the input and output variables. Carefully tuning the hyperparameters through cross-validations reduces the risk of overfitting and helps ensure good generalization abilities of the models (Alpaydin, 2014).

### 4.3 Main findings

All the forecast models tested in this study performed comparably to or better than commercial forecast alternatives.

When comparing the different model types, the two nonlinear models clearly outperformed the linear OLS model. The SVR and MLP models performed similarly, but the SVR won on accuracy by a small margin. The predictive power of the SVR model is not surprising, since SVR models have shown strong performance for heat load forecasting in the past (Idowu et al., 2014; Izadyar et al., 2015).

The performance of the models was highly dependent on which input variables were chosen. Augmenting the weather and lagged heat load variables with generic calendar data about time of day, month of the year etc., lead to significant performance boosts. This is because the generic calendar information helps the models better learn the daily profile of the heat load. The daily heat load profile is a result of consumer behavior and is different on different types of days. Providing the models with information about national holidays, observances and school holidays increased the forecasting accuracy slightly on average. However, the performance increase was most notable on holidays and on weekends.

Incorporating human behavior through the use of local holiday data has potential to improve energy system operations. In district heating systems, operators could expect a better performance of their load forecasts on days that are traditionally difficult to forecast well.

Article

# Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data

Magnus Dahl <sup>1,2,\*</sup> , Adam Brun <sup>2</sup>, Oliver S. Kirsebom <sup>3</sup>  and Gorm B. Andresen <sup>1</sup> 

<sup>1</sup> Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus, Denmark; gba@eng.au.dk

<sup>2</sup> AffaldVarme Aarhus, Municipality of Aarhus, Bautavej 1, 8210 Aarhus, Denmark; adbr@aarhus.dk

<sup>3</sup> Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus, Denmark; oliskir@phys.au.dk

\* Correspondence: magnus.dahl42@gmail.com

Received: 28 May 2018; Accepted: 25 June 2018; Published: 27 June 2018



**Abstract:** The heat load in district heating systems is affected by the weather and by human behavior, and special consumption patterns are observed around holidays. This study employs a top-down approach to heat load forecasting using meteorological data and new untraditional data types such as school holidays. Three different machine learning models are benchmarked for forecasting the aggregated heat load of the large district heating system of Aarhus, Denmark. The models are trained on six years of measured hourly heat load data and a blind year of test data is withheld until the final testing of the forecasting capabilities of the models. In this final test, weather forecasts from the Danish Meteorological Institute are used to measure the performance of the heat load forecasts under realistic operational conditions. We demonstrate models with forecasting performance that can match state-of-the-art commercial software and explore the benefit of including local holiday data to improve forecasting accuracy. The best forecasting performance is achieved with a support vector regression on weather, calendar, and holiday data, yielding a mean absolute percentage error of 6.4% on the 15–38 h horizon. On average, the forecasts could be improved slightly by including local holiday data. On holidays, this performance improvement was more significant.

**Keywords:** district heating; load forecasting; machine learning; weather data; consumer behavior; neural networks; support vector machines

## 1. Introduction

Energy systems are changing throughout the world, and heat load forecasting is gaining importance in modern district heating systems [1]. The growing penetration of renewable energy sources makes energy production fluctuate beyond human control and increases the volatility in electricity markets. Stronger coupling between the heating and electricity sectors means that production planners in systems with combined heat and power generation need accurate heat load forecasts in order to optimize the production.

It is not trivial to forecast district heating demand on time scales that are relevant for trading on the day-ahead electricity market. The total heat load in a district heating system is influenced by several factors—most importantly, the weather, the building mass of the city, and the behavior of the heat consumers. Cold and windy weather increases the heat demand, and warm and sunny weather decreases it. The constitution of the building mass influences how the heat load responds to changes in the weather [2]. Human behavior is an often overlooked factor, and, especially in summer, the heat demand is dominated by hot water consumption rather than space heating. Consumer behavior

can vary considerably from day to day, and the heat load on special occasions, e.g., New Year's Eve, is notoriously difficult to forecast accurately.

Heat load forecasting has been studied extensively in the scientific literature. The successful application of simple linear models in [3,4] has inspired us to use an ordinary least squares (OLS) model as a simple benchmarking model. Statistical time-series models, such as SARIMA (seasonal autoregressive integrated moving average) models [4,5] and grey-box models combining physical insight with statistical modeling [6], are natural ways of handling the temporal nature of load forecasting. These models are usually linear and struggle with multiple seasonality. In [7], the authors compared a number of machine learning algorithms, including a simple feed forward neural network, support vector regression (SVR), and OLS. They concluded that the SVR model performs best. The strong forecasting capabilities of SVR models have also been demonstrated in [8], where heat demand was forecasted based on natural gas consumption. Neural networks have been widely applied in load forecasting. Several studies apply simple feed forward networks with one hidden layer such as the multilayer perceptron (MLP) [7,9,10]. A recurrent neural network is used in [11] to better handle non-stationarities in the heat load. A comprehensive review of load forecasting in districts can be found in [1].

In the present study, we chose to compare three different machine learning models: OLS, MLP, and SVR, as they have all proven effective for heat load forecasting. Some studies attempt to include the different consumer behavior on weekdays and weekends. Working days and non-working days are modeled with distinct profiles in [12], and in [4] mid-week holidays were treated as Saturdays or Sundays. In [13], the correlation between electric load and weather variables was exploited to forecast the aggregated load using MLP models, and the authors explored the different autocorrelations of the load on weekdays and weekends. In this study, we include generic calendar data such as the day of the week, as well as local holiday data to account for observances, national holidays, and city-specific holidays, i.e., school holidays.

School holidays are often planned locally, and some religious holidays, e.g., Easter, fall on different dates each year. Therefore, generic calendar data is insufficient for modeling events that depend on local holidays. Heat consumers behave differently on holidays and change the pattern of consumption, so including local holiday data in heat load forecast models has the potential to improve forecast accuracy.

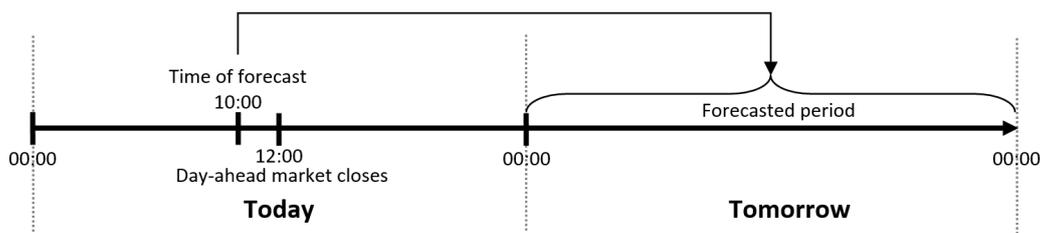
The novelty of this work lies in the application of new data sources, specifically local holiday data, to create heat load forecasting models that more accurately capture consumer behavior. To the best of our knowledge, school holiday data has not previously been used for heat load forecasting. We isolate the effect of using local holiday data by employing machine learning models that have proven effective for heat load forecasting in the past. Moreover, we base our modeling on a very large amount of data. Seven years of hourly heat load and weather data supplemented with data about national holidays, observances, and school holidays help the forecast models capture rare load events.

The remainder of the paper is structured as follows. The Methodology section describes the data foundation, the machine learning models, and the validation and testing procedure. In the Results section, the forecasting models are benchmarked and compared, and the potential of using new data sources is evaluated. The paper is wrapped up in the Conclusion section.

## 2. Methodology

In this section, we describe the data foundation and how the heat load forecasting models were built, validated and tested.

The focus of this paper is to create heat load forecasts that are relevant on the time horizon of the day-ahead electricity market. Therefore, a forecast must be produced each morning at 10:00 for each hour of the following day. This timeline, illustrated in Figure 1, allows time for communication between different actors in a production system and for planning of the following day's heat production in accordance with the bids in the day-ahead electricity market.



**Figure 1.** Timeline for the heat load forecast that is relevant for the trading decisions in the day-ahead electricity market. Every day at 10:00 a forecast is produced for each hour of the following day.

The analysis in this paper is based on seven years of data for the total hourly heat load of Aarhus, Denmark. The years 2009, 2010, 2012, 2013, 2014, 2015, and 2016 were used. Unfortunately, heat load data from 2011 has not been available to us. We denote the heat load in hour  $t$  by  $P_t$ .

### 2.1. Weather Data

The heat demand depends strongly on the weather. Hourly outdoor temperature, wind speed, and solar irradiation for the seven years were obtained from [14]. Weather data from the geographical point N 56°2'42.24", E 9°59'59.95" in the southern part of Aarhus was used. Weather forecasts of the outdoor temperature, wind speed, and solar irradiation were provided by the Danish Meteorological Institute (DMI) and used to test the performance of the heat load forecasts as realistically as possible. These weather forecasts were based on the HIRLAM (High Resolution Limited Area Model), a numerical weather prediction system developed by a consortium of European meteorological institutes with the purpose of providing state-of-the-art short-range weather predictions [15], for numerical weather prediction, had a forecast horizon of up to 54 h, and were disseminated four times a day [15]. We denote the outdoor temperature, wind speed, and solar irradiance by  $T_t^{\text{out}}$ ,  $v_t^{\text{wind}}$  and  $I_t^{\text{sun}}$ , respectively.

### 2.2. Calendar Data

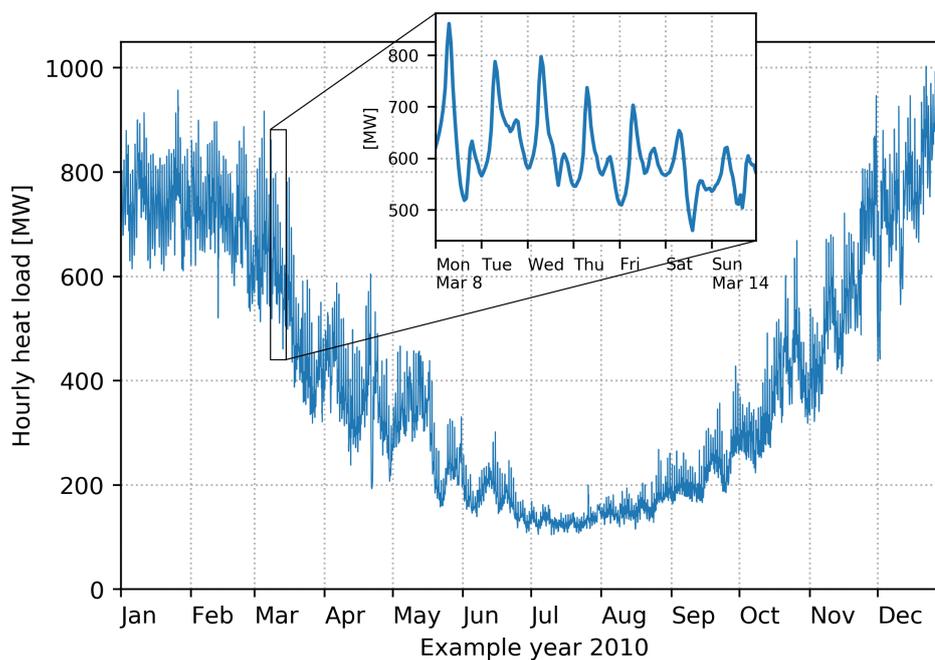
The heat demand has a strong social component that depends on human behavior. The social component is part of the reason for the daily and weekly patterns in the heat load. Different load profiles on weekdays and weekends can also be explained by consumer behavior. In order to allow the forecast models to account for load variations that are tied to specific days, seasons, and times of day, certain calendar data were included as input variables. Specifically, the hour of the day, the day of the week, the weekend, and the month of the year were used as input. How the calendar data was encoded and included in the models is described in Section 2.4.1.

### 2.3. Holiday Data

In addition to generic calendar data, we also used more specific local data about special days that may influence the heat consumption pattern. The district heating system of Aarhus, Denmark, served as our case study. Therefore, we used data about Danish national holidays, observances, and local school holidays. National holidays and observances were sourced from [16]. National holidays include New Year's Day, Christmas Day, Easter Day, etc. and constitute 11 days per year. Observances include, e.g., Christmas Eve and Constitution Day and amount to six days per year. Information about the municipal school holidays was collected from local schools in the Aarhus area and amounts to 96 days per year on average. Note that all national holidays are also school holidays. It is clear that this kind of information is highly local and that gathering such data, compared to the generic calendar data, is more difficult. The following analysis will illuminate whether including this data significantly improves heat load forecasts, or if more easily available data types are sufficient.

#### 2.4. Data Exploration

Figure 2 shows the average hourly heat load for the example year of 2010. Notice how much the heat load varies over the year both in magnitude and in variance. The zoomed inset in the plot shows the heat load variations over a week in March. A clear daily pattern can be observed, with a sharp morning peak between 7:00 and 8:00 on weekday mornings. The morning peak is a well-known phenomenon in the district heating community and is caused by many people showering around the same time every morning. On weekends, morning peaks can be observed later in the morning and tend to be less sharp compared to weekdays. From the inset, it is clear that the daily load pattern varies substantially within just one week.



**Figure 2.** Time series for the hourly heat load in the year 2010. The inset shows a zoom of a week in March.

The heat demand has peaks in its autocorrelation function at 24 h, 48 h, 72 h, and so on. This is due to the strong daily pattern. There is also a notable peak at 168 h (one week). In order to capture this behavior, lagged values of the heat load were used as input variables in the modeling. Specifically, we included the heat load lagged with 24 h, 48 h, and 168 h. Looking at Figure 1, we see that the forecast horizon varies between 15 h and 38 h. The heat load in the first hour of the day can be forecasted with the shortest horizon, and the last hour of each day is forecasted with the longest horizon. When forecasting hours with a forecast horizon of 24 h or less, the heat load lagged 24 h can be used. When forecasting hours with a longer horizon than 24 h, the heat load lagged with 48 h must be used instead. A power spectrum analysis confirmed strong peaks at frequencies  $1/12 \text{ h}^{-1}$  and  $1/24 \text{ h}^{-1}$ , but 12 h is shorter than the shortest forecast horizon and was thus discarded. The two lags that best captured the daily and weekly pattern of the heat load were included. We denote the lagged heat load by  $P_{t-24}$ ,  $P_{t-48}$ , and  $P_{t-168}$ , respectively.

The most important weather variable when modeling district heating loads is the outdoor temperature, because there is a strong negative correlation between the heat demand and the outdoor temperature. Depending on the specific district heating system, solar irradiation and wind speed can also be significant predictors for the heat load [2]. Due to the thermal mass of the buildings in a district heating system, there is a certain inertia in the heat load when changes in the weather occur. On the individual building level, this inertia is handled in great detail in the civil engineering

literature [17]. Since we were forecasting the heat load of an entire city, we took a more simplified approach. In the Aarhus district heating system, the heat load is most strongly correlated with the outdoor temperature lagged by 4 h, compared to other time lags of the temperature. The heat load also correlates most strongly with the solar irradiation lagged by 4 h. There seemed to be no benefit in lagging the wind speed. Only including two specific lags, is of course a simplification of the dynamics of the system, but the results of including them were significantly better than just using simultaneous (lag 0 h) weather variables. Summing up, the following five weather variables were included in the modeling:  $T_t^{\text{out}}$ ,  $v_t^{\text{wind}}$ ,  $I_t^{\text{sun}}$ ,  $T_{t-4}^{\text{out}}$ , and  $I_{t-4}^{\text{sun}}$ .

Outdoor temperature and, as a consequence, the heat load varies substantially from year to year [18]. The mean annual temperatures in our dataset spanned a range of 2.5 °C. Compared to the mean load of the whole dataset (excluding 2016), the annual mean heat load was 15% higher in the coldest year and 11% lower in the warmest year.

#### 2.4.1. Data Scenarios and Pre-Processing

In order to evaluate the effect of including the various types of input data for forecasting heat load, three different data scenarios have been defined. We call these scenarios: “Only Weather Data,” “Weather and Calendar,” and “Weather, Calendar, and Holidays”. Table 1 details the input data used in each scenario.

**Table 1.** Input variables used in the three data scenarios (in bold).

		<b>Only Weather Data</b>	<b>Weather and Calendar</b>	<b>Weather, Calendar and Holidays</b>
Lagged heat load	$P_{t-24}$ or $P_{t-48}$	✓	✓	✓
	$P_{t-168}$	✓	✓	✓
Weather data	$T_t^{\text{out}}$	✓	✓	✓
	$v_t^{\text{wind}}$	✓	✓	✓
	$I_t^{\text{sun}}$	✓	✓	✓
	$T_{t-4}^{\text{out}}$	✓	✓	✓
	$I_{t-4}^{\text{sun}}$	✓	✓	✓
Calendar data	Hour of day		✓	✓
	Day of week		✓	✓
	Weekend		✓	✓
	Month of year		✓	✓
Holiday data	National holiday			✓
	Observance			✓
	School holiday			✓

To achieve the best performance of the models, the input data were scaled and encoded as follows. All the continuous variables (lagged heat load and weather) were standardized to have mean 0 and standard deviation 1. The calendar data and holiday data were included as so-called dummy variables. Dummy variables are a way to represent categorical variables as binary variables. For instance, whether or not a given hour falls on a school holiday can be encoded as a binary variable (0 or 1). The day of week can be encoded as six binary variables: one variable indicating if it is Monday, one indicating if it is Tuesday, etc. Only six variables are needed to encode seven days, because if it is not any of the days from Monday to Saturday, then it must be Sunday. Using similar dummy variables all the calendar and holiday data was included. Encoding categorical data as dummy variables is a standard machine learning method [19].

#### 2.5. Machine Learning Models

We benchmarked and compared three different machine learning models that have all previously been proven adequate for heating load forecasting [7,8]: ordinary least squares regression, multilayer perceptron, and support vector regression.

### 2.5.1. OLS—Ordinary Least Squares Regression

Ordinary least squares regression is a simple model type in which the output is modeled with the hyperplane that minimizes the squared residuals between the target and the output of the model. Sometimes referred to as multiple linear regression, it is a popular model due to its simplicity, computational speed, and the fact that results can be easily interpreted. Because of its linear structure, the OLS model underperforms when modeling nonlinear input–output relationships.

### 2.5.2. MLP—Multilayer Perceptron

A multilayer perceptron is a simple kind of neural network. Neural networks have been applied to problems in many fields, including heat load forecasting, due to their ability to capture complicated relationships between input and output [7,9,11]. A multilayer perceptron has at least one hidden layer between the input and output layers of the model and a nonlinear activation function allows the model to capture nonlinear relationships between input and output. A good coverage of neural network models and the multilayer perceptron can be found in [20]. We used a multilayer perceptron with one hidden layer and the rectifier activation function:  $f(x) = \max(0, x)$ . We have experimented with adding more hidden layers, but the increase in the model accuracy was not large enough to justify the growth in model complexity and the risk of overfitting.

Besides the simple multilayer perceptron, we have experimented with a more advanced type of neural network. Recurrent neural networks with long short-term memory (LSTM) units [21] were implemented in an attempt to simplify the feature selection and leave it to the model to discover relevant lags of heat load and weather data. Our initial LSTM networks yielded results comparable to the simpler models included in this work, but their performance tended to be inconsistent. The benefit of simplified feature selection may also be outweighed by a more complicated model selection and training procedure. The LSTM modeling for heat load forecasting requires more work and will be left for future work.

### 2.5.3. SVR—Support Vector Regression

Support vector regression is the application of support vector machine models to regression problems and was first introduced in [22]. Support vector regression has a computational advantage in very high dimensional feature spaces. The model only depends on a subset of input data, because it minimizes a cost function that is insensitive to points within a certain distance from the prediction. The cost function is less sensitive to small errors and less sensitive to very large errors and outliers, compared to the quadratic cost function minimized in the ordinary least squares regression. To avoid overfitting, the model is governed by a regularization parameter  $C$ , that ensures that the parameters of the model do not grow uncontrollably. The smaller the value of  $C$ , the harder large model parameters are penalized. Support vector regression is explained in great detail in [19,20]. By employing the so-called “kernel trick”, support vector regression can handle nonlinear input–output relationships. A very popular kernel function is the radial basis function kernel (RBF), which has been proven effective in this application as well. The RBF kernel is governed by a kernel parameter  $\gamma$ . The greater the value of  $\gamma$ , the more prone the model is to overfitting, but if  $\gamma$  is chosen too small, the model may be underfitting and fail to capture actual input–output relationships.

Summing up, the three machine learning models OLS, MLP, and SVR were chosen because they have all been successfully applied to heat load forecasting in the past. Using well-established algorithms allows us to focus on the main research question: whether conventional heat load forecasts can be improved by adding new types of data. Each of the models have advantages and disadvantages. The advantage of the OLS model is that it is computationally cheap, and its estimated parameters carry a physical interpretation. The disadvantage is that the model is linear and fails to capture nonlinearities in input–output relationships. The advantage of the MLP model is that it is capable of capturing very complex relationships between input and output. A disadvantage of neural network models, such as

MLP, is the risk of overfitting and that they require careful tuning of several hyperparameters. Finally, the SVR model has the advantage of being robust to outliers and that the final model depends only on a subset of the training data. The SVR model, however, is sensitive to the scaling of the input data and the correct tuning of regularization and kernel parameters.

### 2.6. Model Selection and Testing

A good forecast model is one that performs well on previously unseen data. This is the generalization ability of the model. In order to accurately measure the generalization performance of the models, we divided the full dataset (seven years of hourly data) into a training and validation set and a test set. All model selection and training was performed on the years from 2009 to 2015 (2011 not included). This is the training and validation set. The entire year of 2016 was used as a blind test set to estimate the generalization performance of the forecasts.

The three models were chosen and their hyperparameters tuned based on sixfold cross-validation on the years 2009, 2010, 2012, 2013, 2014, and 2015. Using six folds ensured that each fold contained an entire year and thus represented the full annual variation of the heat load. In the cross-validation, the different models and data scenarios were scored according to the hourly root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_t (\hat{P}_t - P_t)^2} \quad (1)$$

where  $\hat{P}_t$  is the forecasted heat load for hour  $t$ , and  $N$  is the number of hours.

The OLS model does not have any hyperparameters to tune, but a model with a nonzero constant term was chosen. In the MLP model, we tuned the number of neurons in the hidden layer using a grid search on the cross-validation scores. A MLP model with one hidden layer consisting of 110 hidden neurons was chosen, and the L2 regularization parameter  $\alpha$  was set to 0.1. In the SVR model, the best choices for the regularization parameter and the kernel parameter were found to be  $C = 4.3$  and  $\gamma = 0.2$ . All modeling has been performed in Python 2.7 using the scikit-learn framework (version 0.19.0) [23].

All results presented in the following section were produced using the blind test year 2016. This year was not used for any of the training, data exploration, or model selection. In the Results section, we employ two other forecast error metrics, in addition to the RMSE. The mean absolute error (MAE) is also an absolute error metric (here in units of MW), but it is less sensitive to large errors, compared to the RMSE. The MAE is defined as

$$\text{MAE} = \frac{1}{N} \sum_t |\hat{P}_t - P_t| \quad (2)$$

Finally, we use the relative error metric mean absolute percentage error (MAPE) to facilitate easier comparison between different district heating systems. The MAPE is defined as

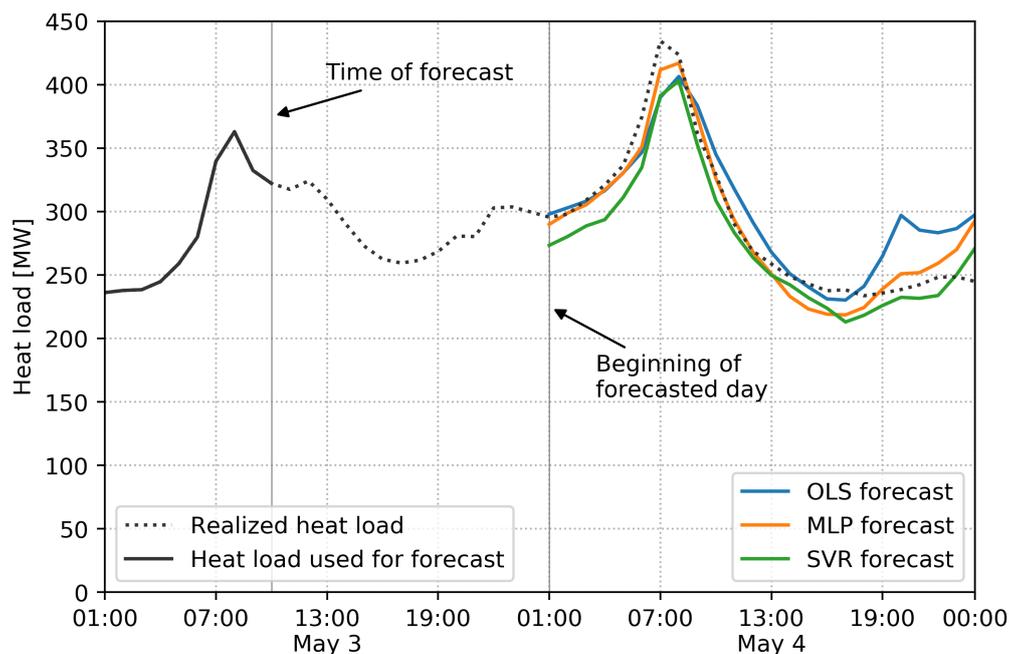
$$\text{MAPE} = \frac{1}{N} \sum_t \left| \frac{\hat{P}_t - P_t}{P_t} \right| \quad (3)$$

## 3. Results

The heat load in a district heating system has been forecasted using three different machine learning models, described in the previous section: OLS, MLP, and SVR. The performance of these models have been tested by letting them produce a forecast for the following day using the input data available each day at 10:00 a.m. The models have been trained exclusively on data prior to the test year 2016 to be able to accurately gauge their generalization performance. Figure 3 shows an example of

the forecasts produced for 4 May. Only the heat load up to the time of the forecast was used as input to produce the forecast. Real weather forecasts were used as weather inputs for 4 May, as opposed to the historical weather data used for training. It is clear how the three forecast models produce similar, yet distinct forecasts. On 4 May, the MLP model appears to produce the best forecast, especially in the morning.

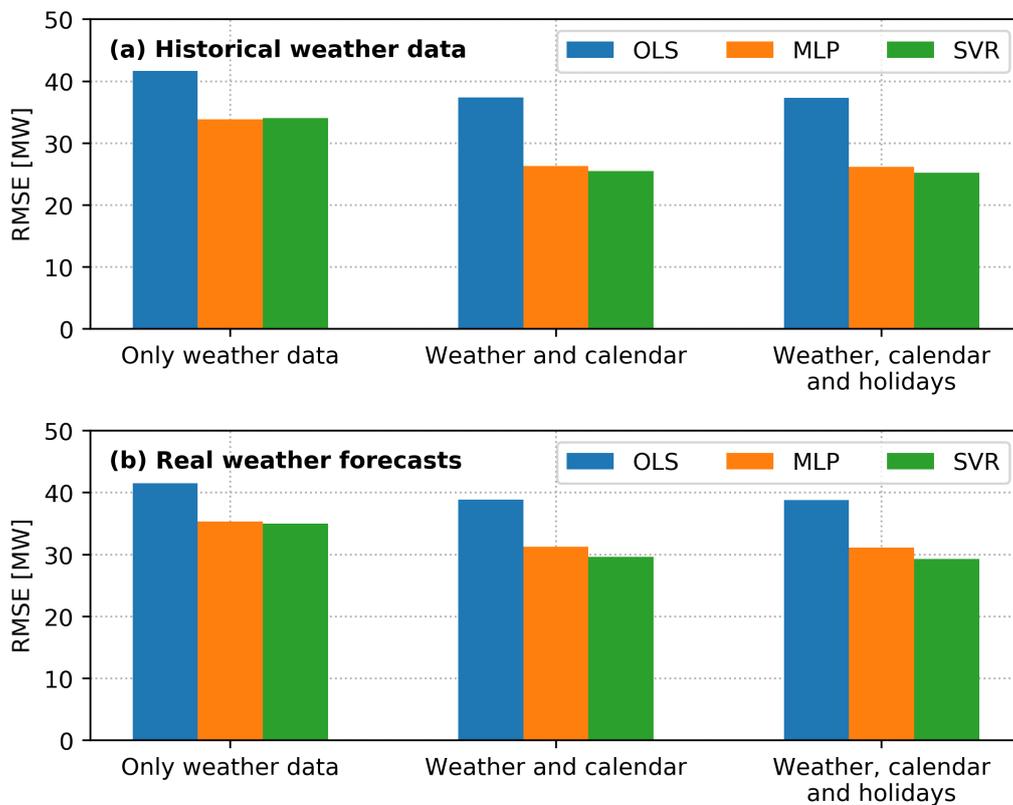
Figure 4 summarizes the performance of the three models in the three different data scenarios. The top panel shows the forecast performance that could be achieved if weather forecasts were 100% accurate, simulated by using historical weather data. The bottom panel shows the performance using real weather forecasts. Comparing the three data scenarios, we see the benefit of including different data types in the modeling. In the first scenario, only lagged heat load and weather data are used as input. In the second scenario, generic calendar data is included as well, and in the third scenario, local observances, national holidays, and school holidays are also included as inputs to the model. Including calendar data significantly improves performance, compared to only using weather data. Extending the input data with holiday data as well results in an additional, but small improvement compared to using generic calendar data only. Obtaining the local holiday data can be laborious or impossible, so it is positive to see generic calendar data yielding comparable results. It is much easier to apply these models to a wide range of district heating systems around the world if it can be done without collecting local holiday data.



**Figure 3.** Example forecasts for 4 May 2016. The forecasts were produced on 3 May at 10:00 and based on real weather forecasts, calendar, and holiday data.

Figure 4 allows for comparison of the performance of the three machine learning models as well. The OLS model stands out by performing significantly worse than the other two models in all scenarios. The OLS model has a root mean square error of 38.9 MW, compared to 31.1 MW and 29.3 MW for the other two models when using real weather forecasts, calendar, and holiday data (bottom panel). The poor performance of the OLS model can be attributed to its linear structure. The relationship between the outdoor temperature and the heat load in a temperate climate is nonlinear. This causes the linear model to perform poorly during summer by undershooting the heat load and overestimating its variance. The two nonlinear models, MLP and SVR, perform similarly in these scenarios. The SVR

model has the smallest error, and the focus in the rest of this paper will be on the SVR model using weather, calendar, and holiday data.



**Figure 4.** Root mean square error of the three forecast models OLS, MLP, and SVR on the year 2016. The top panel (a) shows the error using historical weather data to simulate 100% accurate weather forecasts. The bottom panel (b) shows the error using real weather forecasts.

### 3.1. The Value of Improving Weather Forecasts

Figure 4 has two panels. The top panel shows the forecast errors that could be achieved if weather forecasts predicted the measured weather completely accurately. This has been simulated by allowing the models to use actual measured weather data, instead of weather forecasts as input when producing the load forecast. The top panel reflects the scenario in the which future weather is known. The bottom panel shows the results in the case where real forecast data has been used instead. This is the actual forecast performance that can be achieved in an operational situation, given the current quality of weather forecasts. Having access to weather forecasts without prediction errors could, in a perfect world, reduce the error from 29.3 MW to 25.2 MW in the forecasts from the best model. While an error reduction of 4.1 MW is a start, perfecting the weather forecast only shaves 14% off the error. The remainder of the load forecast error has other causes than weather forecast errors, a result that was also found in [24], where ensemble weather predictions were used to quantify heat load forecasting uncertainty.

The OLS model using only weather data and lagged heat load does not perform notably different on historical weather data compared to real weather forecasts. This can be explained by the OLS model attributing greater weight to the lagged heat load compared to the weather, because the relationship between the heat load and the weather is nonlinear.

The forecast performance, shown in the top of Figure 4, is similar to the performance that was achieved during training and cross-validation. This indicates that the models have not been overfitted and generalize well to out-of-sample predictions.

It is worth pointing out that the performance of all these models, even the OLS model using only weather data, exceeds the performance of the commercial forecasting system that is currently in operation in the Aarhus district heating system. This commercial forecasting system had an RMSE of 41.9 MW in year of 2016 on the same forecast horizons. In relative terms, the SVR model has a MAPE of 6.4% versus 8.3% for the commercial system. The models presented here perform better than all other forecast models that have been used in the Aarhus district heating system.

### 3.2. Seasonal Performance Variations

The heat load varies significantly over the year, both in magnitude and in variance, as exemplified in Figure 2. It can be a challenge for a single model to adequately forecast both winter and summer heat loads. Therefore, it is relevant to further investigate the model performance throughout the year. The forecast error of the best model, SVR using weather, calendar, and holiday data, is illustrated in Figure 5. Three different error metrics are shown: on the left axes, the RMSE (blue) and MAE (yellow) are shown in MW; on the right axes, the MAPE (red) is shown in percent. The horizontal axes show the hour of day for the forecasted hour, and each subplot depicts a month in the year. This makes it possible to see if it is harder to forecast the morning peak and if the forecast horizon impacts the accuracy. Keep in mind that Hour 1 has the shortest forecast horizon (15 h), and Hour 24 has the longest horizon (38 h), since the forecasts are produced at 10:00 a.m. the previous day.

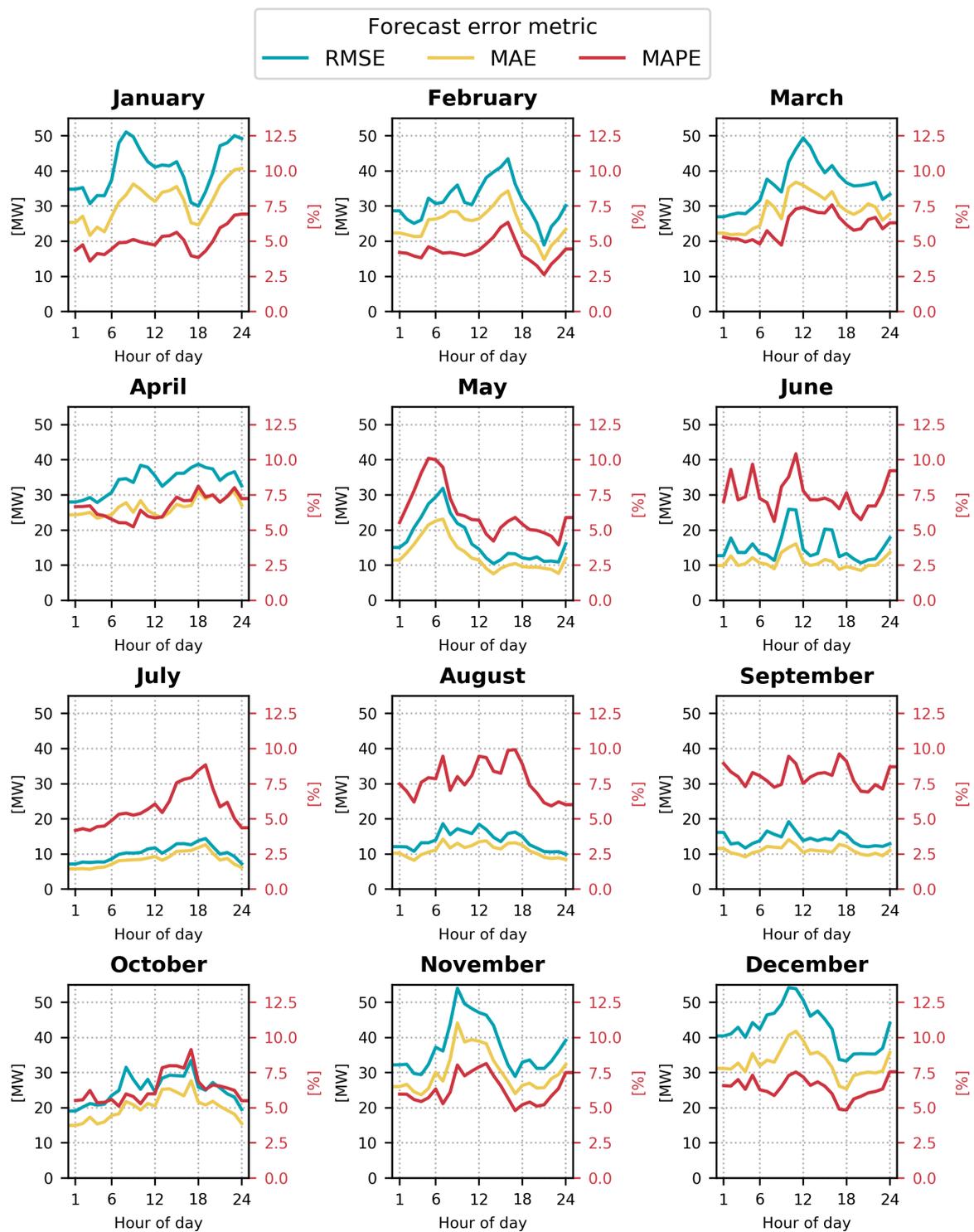
Inspecting Figure 5, it is clear that the absolute error measures RMSE and MAE are largest in winter and smallest in summer. This is a reflection of the annual heat load profile and the large load with large variance during winter. In late fall and winter, the RMSE can be above 50 MW in some hours, whereas it can be below 10 MW in some hours in July. The relative error metric MAPE behaves in the opposite way. The relative error is smaller in the winter months and larger in summer months, but it stays between 2.5% and 10.5%. This is a consequence of the annual load variations being larger than the annual variations in the absolute error.

There is no clear pattern in the way the error changes during the day. The model does not seem to perform worse between 7:00 and 8:00 in the morning, where the morning peak falls. November and May are exceptions to this rule. In many applications, the error of a forecast model increases with the forecast horizon (here the hour of day). We do not observe a general increasing trend in the error with the hour of day. This indicates that the weather forecasts that are used as inputs to create the forecast are not significantly worse at the longest horizon compared to the shortest horizon. It may also be due to weather forecasting accuracy having a minor impact on the heat load forecasting error, as we saw from Figure 4. If we were to increase the forecast horizon further, the forecast error would most likely increase.

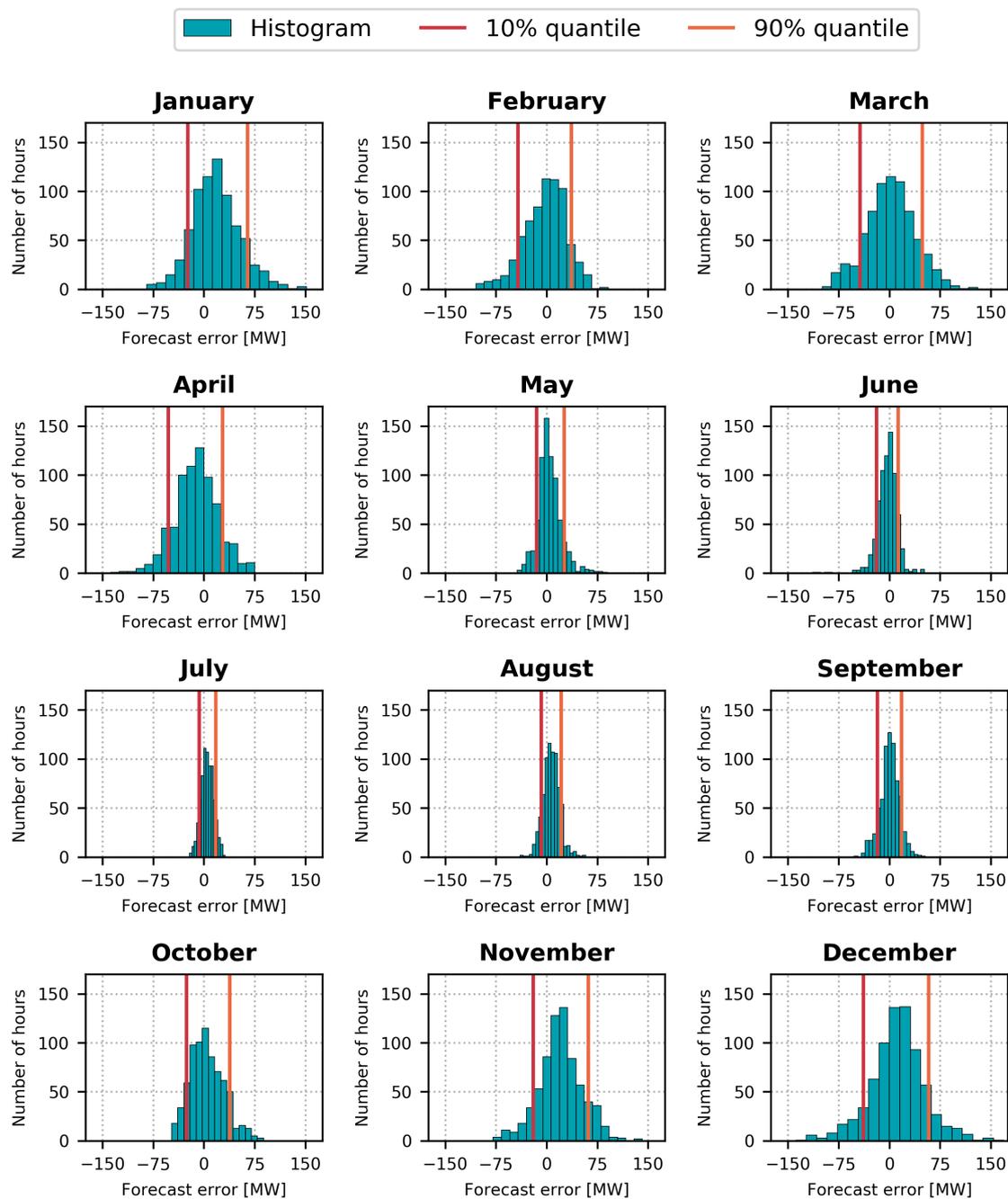
The forecast error varies significantly over the year, but aggregated error metrics such as RMSE, MAE, or MAPE do not tell the full story. Maximum errors can be relevant for unit commitment in the production planning and for evaluating risk regarding trading in the electricity market. Figure 6 shows histograms for the hourly error for each month of the blind test year 2016. The 10% and 90% quantiles have been indicated in each plot. It is clear that the width of the error distribution varies substantially from month to month. During the summer, the forecast error is quite confined, but the distribution widens in late fall and becomes widest in December.

In Table 2, a summary of the error distribution is shown. The 99% and 1% quantiles of the error distribution are indications of the maximum errors that can be expected. Ninety-eight percent of the forecasted hours have forecast errors between the 1% and the 99% quantile. The best month is July with 98% of the errors falling between  $-16.0$  and  $25.8$  MW. The worst month is December, where there is a 1% risk of the forecast overshooting by more than  $115.0$  MW and a 1% risk of the forecast

being more than 96.7 MW too low. These extreme errors can approach 20% of the mean heat load in December.



**Figure 5.** Performance of the SVR model on the year 2016, using real weather forecasts, calendar, and holiday data. Three different error metrics are shown for each month of the year. The forecast error varies with the time of day, shown on the horizontal axes. RMSE (blue) and MAE (yellow) are shown units of MW on the left axes. MAPE (red) is shown in percent on the right axes.



**Figure 6.** Histograms for the forecast error of the SVR model on the year 2016 using real weather forecasts, calendar, and holiday data. The distribution of the forecast error is depicted for each month in the year along with the 10% and 90% quantiles. The number of bins was chosen using Scott's rule [25] within each month. A positive error indicates that the forecast was too high, a negative error that it was too low.

From the histograms in Figure 6, it is also clear that the error distributions are not completely symmetric around 0. In January, for instance, the distribution is shifted slightly to the positive, and in April it is shifted to the negative side. The forecast appears to be biased differently in different months. The mean error for each month (ME) is shown in Table 2. The bias can be as large as 20.5 MW,

with November being the worst month. September performs the best with a mean error of merely 0.3 MW. Varying monthly biases could be remedied by training separate models for each month. However, the focus in this paper is to investigate the effects of including holiday data to model human behavior, and training monthly forecast models would obscure the effects of using holiday data. There is also the possibility that the weather forecasts perform differently at different times of year.

**Table 2.** Summary of the hourly forecast error for each month for the SVR model using real weather forecasts, calendar, and holiday data. Histograms of the forecast error appear in Figure 6. The months with the worst performance are indicated in red, the best in green. The quantiles are evaluated in pairs, so the widest symmetric quantile interval is considered the worst.

	RMSE (MW)	ME (MW)	Error Quantiles (MW)			
			10%	90%	1%	99%
January	41.2	18.7	-24.0	64.1	-66.8	117.6
February	31.8	-2.2	-42.8	36.6	-91.6	61.3
March	36.9	2.0	-43.7	48.8	-80.4	89.3
April	34.2	-11.7	-52.9	27.1	-93.6	64.8
May	18.2	4.3	-14.7	25.6	-33.6	64.3
June	15.8	-3.3	-19.3	12.8	-45.3	34.0
July	10.6	5.0	-7.0	17.2	-16.0	25.8
August	14.2	6.9	-8.2	21.6	-20.1	41.6
September	14.3	0.3	-17.9	17.6	-35.0	33.3
October	25.6	4.8	-26.0	37.8	-43.5	70.0
November	38.1	20.5	-20.0	61.5	-59.1	98.1
December	43.1	12.5	-38.4	58.2	-96.7	115.0

In conclusion, there are significant seasonal variations in the performance of the best heat load forecast. The absolute errors are largest in winter and smallest in summer, with December being the hardest month to forecast and July being the easiest.

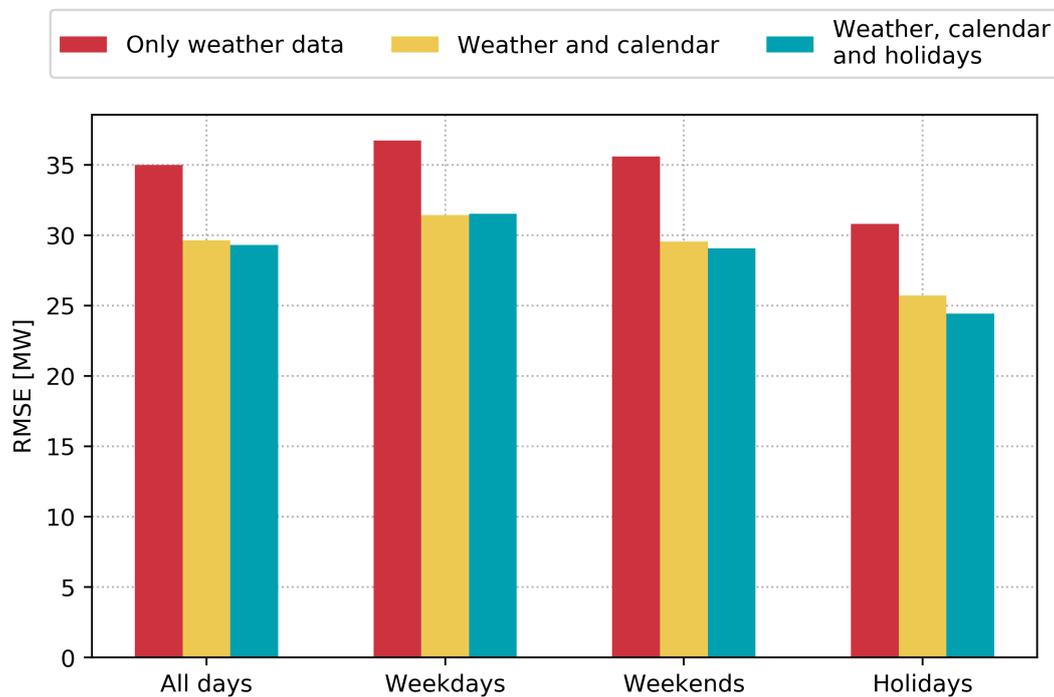
### 3.3. The Value of Calendar and Holiday Data

The goal of this analysis is to gauge the potential of including local holiday data in heat load forecasts in order to better capture the consumer behavior. The reduction in the annual error was very small when comparing models with only generic calendar data to models including local holiday data. This was clear from Figure 4b. It is well known among district heating operators that heat load forecasts tend to perform poorly on special occasions, such as Christmas or New Year's Eve. These special days are rare, so the performance on those specific days has little impact on the average annual performance (Figure 4b). Improved performance on special days is valuable to production planners, and whether including local holiday data can improve forecast performance on specific days is worth investigating in more detail.

Figure 7 shows the performance of the SVR model in the three data scenarios on different sets of days during the year. "Holidays" refer to all days that are observances, national holidays, or school holidays. "Weekdays" include all weekdays that are not also in holidays, and "weekends" include all weekend days not included in holidays. In 2016, there were 201 weekdays, 65 weekend days, and 100 holidays.

There is significant benefit in including generic calendar data in the forecast models for all day types. On weekdays, there is no performance improvement to gain by including local holiday data. The forecast error on weekends can be reduced by 0.5 MW. Not surprisingly, the greatest performance increase can be observed on holidays. The holiday error decreases by 1.3 MW when augmenting the modeling with local holiday data. The holiday error is generally smaller than the error for the other day types. This is due to the holidays being dominated by the schools' summer holidays, and the error is generally smaller during the summer. Summing up, including local holiday data only improves

the forecasts slightly on average. The largest improvement is seen on holidays where the error can be reduced by 5%, compared to only using generic calendar data.



**Figure 7.** Forecast performance of the SVR model on the year 2016 using real weather forecasts, calendar, and holiday data. The second and third group of bins refer to weekdays and weekends that are not also included in holidays. Holidays refer to all days that are observances, national holidays, or school holidays (see Table 1).

#### 4. Conclusions

We have tested heat load forecasts with horizons from 15 h to 38 h, relevant for district heating production planning considering the day-ahead electricity market. The work was based on seven years of heat load and weather data for the large district heating system of Aarhus, Denmark. In order to measure the forecast performance that can realistically be experienced in actual operation, we used blind testing on a whole year with real weather forecasts.

Three machine learning models have been tested: an ordinary least squares model, a multilayer perceptron, and a support vector regression model. The SVR model performed best, beating the OLS model by a large margin and the MLP model by a small margin. All the models were trained on lagged heat load data and weather data. The forecast performance could be significantly improved by including generic calendar data, such as month, weekday, and hour of day. A smaller improvement of the forecasts could be gained by supplying the models with local holiday data including observances, national holidays, and school holidays. This improvement was most significant on holidays and weekends. Local holiday data can be difficult and time-consuming to obtain, but merely including lagged heat load, weather, and generic calendar data can provide a good overall forecast performance.

The SVR model using weather, calendar, and holiday data had the best performance. The root mean square error was 29.3 MW, and the mean absolute percentage error was 6.4%. This forecast model beat all other models that we have seen for the Aarhus system. The commercial forecast system, currently in operation in the Aarhus district heating system, had an RMSE of 41.9 MW, and a MAPE of 8.3% on the test year.

Including local holiday data showed only minor overall improvements in forecast performance, and including new data types in forecast models requires a careful evaluation of the trade-off between

forecast accuracy and reliability of the data source. In live operational forecast systems, reliability is valued highly, and inputting data into a simpler model may work to make a more robust system. More features are thus not always an advantage, if the improvement in accuracy is insufficient to justify the added implementation and maintenance cost.

Initial experiments using long short-term memory networks have not shown notable improvement over the results attainable with the SVR model. However, future works should explore this type of model further, as it has the potential to simplify the feature selection procedure and make it easier to transfer these results to a wide range of district heating systems around the world.

**Author Contributions:** Conceptualization, G.B.A. and M.D.; Methodology, G.B.A. and M.D.; Formal Analysis, M.D.; Investigation, M.D. and O.S.K.; Data Curation, M.D.; Writing—Original Draft Preparation, M.D.; Writing—Review & Editing, G.B.A. and M.D.; Visualization, M.D.; Supervision, G.B.A. and A.B.; Project Administration, A.B.; Funding Acquisition, G.B.A. and A.B.

**Funding:** This research has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no ENER/FP7/609127/READY.

**Acknowledgments:** We would like to thank the Danish Meteorological Institute for providing the weather forecast data. We also thank AffaldVarme Aarhus for providing data about the heat load and production system in Aarhus.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

$P_t$	Heat load in hour $t$ (MW)
$P_{t-l}$	Heat load lagged by $l$ hours (MW)
$T_t^{\text{out}}$	Outdoor temperature in hour $t$ ( $^{\circ}\text{C}$ )
$v_t^{\text{wind}}$	Wind speed in hour $t$ (m/s)
$I_t^{\text{sun}}$	Solar irradiation in hour $t$ ( $\text{W}/\text{m}^2$ )
$T_{t-l}^{\text{out}}$	Outdoor temperature lagged by $l$ hours ( $^{\circ}\text{C}$ )
$I_{t-l}^{\text{sun}}$	Solar irradiation lagged by $l$ hours ( $\text{W}/\text{m}^2$ )
$\hat{P}_t$	Heat load forecasted for hour $t$ (MW)
$\alpha$	L2 regularization parameter of the MLP model
$C$	Regularization parameter of the SVR model
$\gamma$	RBF kernel parameter of the SVR model
RMSE	Root mean square error (MW)
MAE	Mean absolute error (MW)
MAPE	Mean absolute percentage error (%)
ME	Mean error (MW)
OLS	Ordinary least squares regression model
MLP	Multilayer perceptron model
SVR	Support vector regression model
RBF	Radial basis function kernel
LSTM	Long short-term memory network model

## References

1. Ma, W.; Fang, S.; Liu, G.; Zhou, R. Modeling of district load forecasting for distributed energy system. *Appl. Energy* **2017**, *204*, 181–205, doi:10.1016/j.apenergy.2017.07.009. [CrossRef]
2. Frederiksen, S.; Werner, S. *District Heating and Cooling*; Studentlitteratur: Lund, Sweden, 2013.
3. Dotzauer, E. Simple model for prediction of loads in district-heating systems. *Appl. Energy* **2002**, *73*, 277–284, doi:10.1016/s0306-2619(02)00078-8. [CrossRef]
4. Fang, T.; Lahdelma, R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy* **2016**, *179*, 544–552, doi:10.1016/j.apenergy.2016.06.133. [CrossRef]

5. Grosswindhager, S.; Voigt, A.; Kozek, M. Online Short-Term Forecast of System Heat Load in District Heating Networks. In Proceedings of the 31st International Symposium on forecasting, Prague, Czech Republic, 26–29 June 2011.
6. Nielsen, H.A.; Madsen, H. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy Build.* **2006**, *38*, 63–71, doi:10.1016/j.enbuild.2005.05.002. [CrossRef]
7. Idowu, S.; Saguna, S.; Åhlund, C.; Schelén, O. Forecasting heat load for smart district heating systems: A machine learning approach. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014.
8. Izadyar, N.; Ghadamian, H.; Ong, H.C.; Moghadam, Z.; Tong, C.W.; Shamshirband, S. Appraisal of the support vector machine to forecast residential heating demand for the District Heating System based on the monthly overall natural gas consumption. *Energy* **2015**, *93*, 1558–1567, doi:10.1016/j.energy.2015.10.015. [CrossRef]
9. Kusiak, A.; Li, M.; Zhang, Z. A data-driven approach for steam load prediction in buildings. *Appl. Energy* **2010**, *87*, 925–933, doi:10.1016/j.apenergy.2009.09.004. [CrossRef]
10. Powell, K.M.; Sriprasad, A.; Cole, W.J.; Edgar, T.F. Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy* **2014**, *74*, 877–885, doi:10.1016/j.energy.2014.07.064. [CrossRef]
11. Kato, K.; Sakawa, M.; Ishimaru, K.; Ushiro, S.; Shibano, T. Heat load prediction through recurrent neural network in district heating and cooling systems. In Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics, Singapore, 15–16 May 2008; pp. 1401–1406.
12. Nielsen, T.S.; Madsen, H. Control of Supply Temperature in District Heating Systems. In Proceedings of the 8th International Symposium on District Heating and Cooling, Trondheim, Norway, 14–16 August 2002.
13. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Calavia, L.; Carro, B.; Sánchez-Esguevillas, A.; García, P.; Lloret, J. Experimental analysis of the input variables' relevance to forecast next day's aggregated electric demand using neural networks. *Energies* **2013**, *6*, 2927–2948. [CrossRef]
14. Saha, S.; Moorthi, S.; Pan, H.L.; Wu, X.; Wang, J.; Nadiga, S.; Tripp, P.; Kistler, R.; Woollen, J.; Behringer, D.; et al. The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1015–1057, doi:10.1175/2010BAMS3001.1. [CrossRef]
15. Unden, P.; Rontu, L.; Järvinen, H.; Lynch, P.; Calvo, J.; Cats, G.; Cuxart, J.; Eerola, K.; Fortelius, C.; Garcia-Moya, J.A.; et al. *HIRLAM-5 Scientific Documentation*; Technical Report; Swedish Meteorological and Hydrological Institute: Norrköping, Sweden, 2002.
16. Holidays in Denmark. Available online: [www.timeanddate.com/holidays/denmark/](http://www.timeanddate.com/holidays/denmark/) (accessed on 13 June 2017).
17. Crawley, D.B.; Hand, J.W.; Kummert, M.; Griffith, B.T. Contrasting the capabilities of building energy performance simulation programs. *BUILD. ENVIRON.* **2008**, *43*, 661–673, doi:10.1016/j.buildenv.2006.10.027. [CrossRef]
18. Dahl, M.; Brun, A.; Andresen, G.B. Decision rules for economic summer-shutdown of production units in large district heating systems. *Appl. Energy* **2017**, *208C*, 1128–1138, doi:10.1016/j.apenergy.2017.09.040. [CrossRef]
19. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2014.
20. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
22. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161.
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
24. Dahl, M.; Brun, A.; Andresen, G.B. Using ensemble weather predictions in district heating operation and load forecasting. *Appl. Energy* **2017**, *193*, 455–465, doi:10.1016/j.apenergy.2017.02.066. [CrossRef]
25. Scott, D.W. On optimal and data-based histograms. *Biometrika* **1979**, *66*, 605–610. [CrossRef]





## Chapter 5

# Seasonal production planning

### 5.1 Motivation

The previous two chapters were focused on the short-term operation of district heating systems, and it is not without reason that the 1-48 h horizon has received a lot of attention in the research. On the short term, it is possible to make accurate predictions that can directly benefit operators, production planners and traders. Seasonal production planning, with a horizon of several months is an entirely different game. Due to larger uncertainties, the value of traditional production planning models dwindles on longer time horizons (Dotzauer, 2003). In the present study, I handle the uncertainties in a seasonal production planning problem by using an extensive amount of weather data.

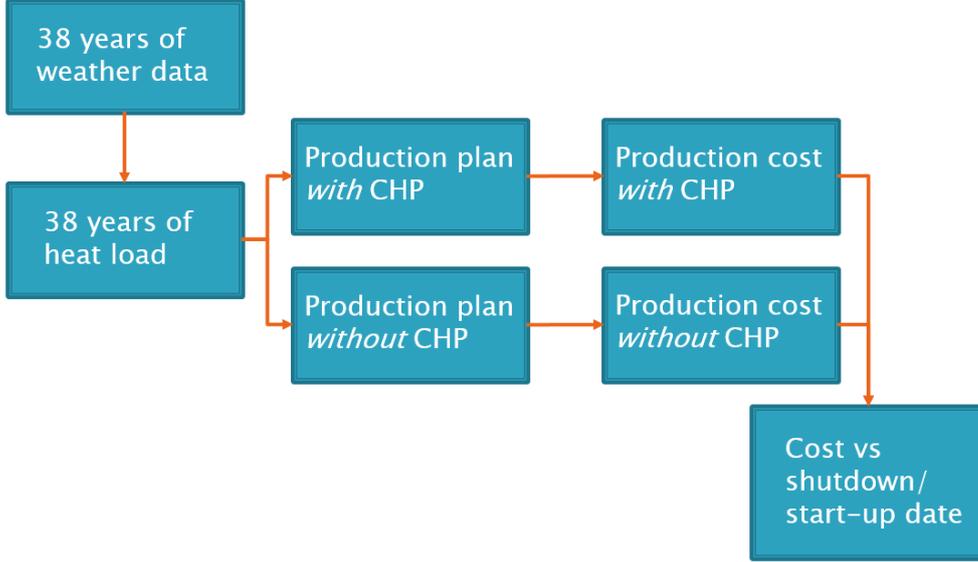
Compared to electric load, the load in district heating systems varies substantially between summer and winter. The seasonal load variations are so large that some heat production units can become superfluous during summer operations in large district heating systems. Shutting down a production unit over the summer has significant economic potential, but the decision is not without risk. If a unit is shut down too early, the system may run into capacity problems, whereas a late shutdown can be costly if it displaces heat from cheaper production units. The optimal shutdown date in the spring and start-up date in the fall vary significantly from year to year and depend on complex weather patterns. Rare weather events, such as an unusually cold spring, can impact a shutdown decision with severe economic risk. Good estimates of this planning risk requires large amounts of weather data, and this study demonstrates that it can be quite costly to make seasonal planning decisions based on only a few years of data.

The main part of this chapter is an article that was published in the Elsevier journal *Applied Energy*. The article demonstrates the benefit of using large amounts of weather data when making seasonal planning decisions in energy systems.

### 5.2 Methods

The core of the study is a simulation model that estimates the heat production cost given a certain heat load and estimates the heat load given weather data.

Based on this simulation model, I proceeded in two parts. First, the variability of the timing of the shutdown and start-up were characterized. Then I developed and benchmarked decision rules that can help production planners salvage most of the potential savings in performing summer shutdown.



**Figure 5.1:** Data flow in the weather-based heat production simulation.

## Heat production simulation

The data flow in the simulation is illustrated in the flowchart in Figure 5.1. In the first step, a model was constructed to calculate the heat load given the weather. Since the purpose of this model was simulation and not forecasting, simplicity was valued highly. My heat load model exploits the well-known hockey stick shape of the heat load plotted against the outside temperature (Frederiksen and Werner, 2013). Daily average values of the heat load were modeled from daily average temperature and an intraday load profile was applied depending on season and whether it was a weekday or weekend. The model is a generalization of piecewise linear degree-day-dependent models and reads:

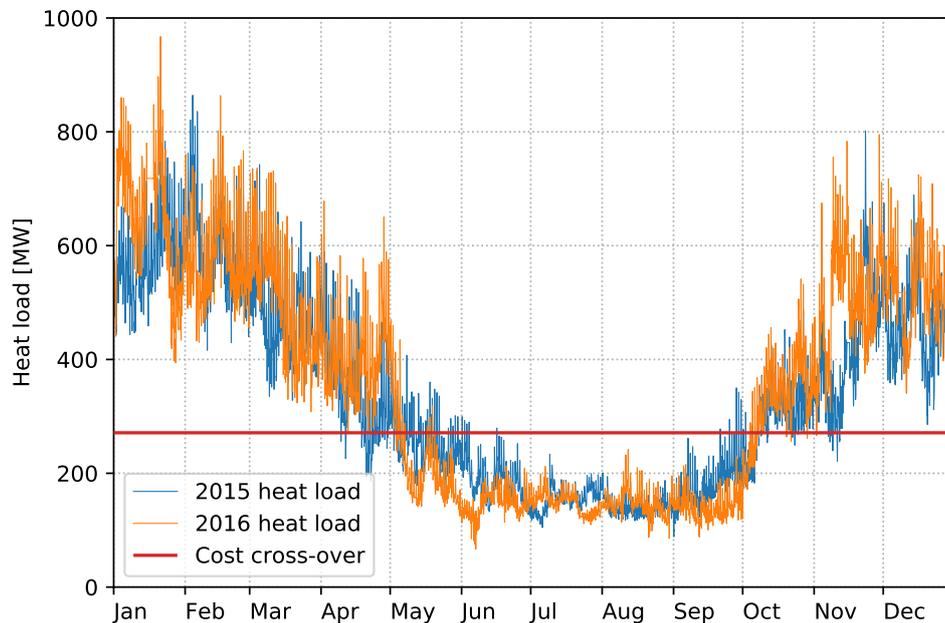
$$P(T_{\text{out}}) = a \left[ (T_{\text{out}} - T_0) \frac{1}{2} \operatorname{erfc} \left( \frac{T_{\text{out}} - T_0}{\sqrt{2}\sigma} \right) - \frac{\sigma}{\sqrt{2\pi}} \exp \left( -\frac{(T_{\text{out}} - T_0)^2}{2\sigma^2} \right) \right] + P_0 . \quad (5.1)$$

Here  $P$  is the heat load and  $T_{\text{out}}$  the outside temperature.  $a$ ,  $T_0$ ,  $\sigma$  and  $P_0$  are parameters in the model. The model was derived under the assumption that the outside temperature, at which the model goes from linear to constant, is normally distributed with mean  $T_0$  and variance  $\sigma^2$ . This model better captures the hockey stick shape of the heat load versus the outside temperature, compared to the piecewise linear models that are often employed for heat load modeling (Dotzauer, 2002; Frederiksen and Werner, 2013).

The next step in the simulation (see Figure 5.1) was to create production plans given a certain heat load. An optimal planning scheme was used to decide which units to run on what capacity in order to cover the heat demand. The optimal production plan was found by solving a mixed integer linear programming problem constrained by the energy balance in the system. Since no heat storage was included in the simulation, this effectively reduced to a simple merit order calculation. Two different production plans were laid: one with and one without the plant scheduled for summer-shutdown.

The cost estimates for the production simulation models were based on the production system of our study case, the Aarhus district heating system. By calculating the production cost with and without the shutdown unit I found the total heat production cost for all possible shutdown and start-up dates. This was the final step of Figure 5.1.

As illustrated on Figure 5.2 the constitution of the production system and the cost of the various production units resulted in a cost cross-over in the heat load. Below the cross-over it was cheaper to shut down the plant, and above it was cheaper to keep the plant running. Shutdown decisions would be easy to make, if the heat load only crossed the cross-over line in one point in the spring and one point in the fall. Notice also the differences between the heat load in 2015 and 2016. The complicated behavior of the heat load around the cost cross-over warrants further studies into optimal shutdown decisions.



**Figure 5.2:** The hourly heat load for Aarhus in 2015 and 2016. Above the red line, it is cheaper to have all production units available. Below the red line, it is cheaper to shut down a large CHP plant.

## Characterization of optimal shutdown

With the production simulation in place, it was possible to begin to answer a first important research question. How much do the cost-optimal shutdown and start-up dates change from year to year, depending on seasonal weather patterns? In addition to the location of the optimum, I also characterized the depth and width of the cost function around the minimum. The depth of the minimum reflects the potential economic savings of performing accurate summer-shutdown. The width indicates how difficult it would be for production planners to realize those savings.

## Decision rules

In order to be able to realize the potential of summer-shutdown, it is not enough to know the past and to know what could have been salvaged given perfect knowledge of the past, present and future. Production planners, who wish to switch between summer and winter operations, need decision rules to tell them the best time to perform shutdown in the spring and start-up in the fall.

In the final part of the study, I developed and benchmarked three different decision rules. The first rule simply shut down the plant on a fixed date every year. The second rule used information about the heat load up to the time of the decision to facilitate more accurate shutdown, but at the cost of shortening the planning horizon. The third rule augmented the second rule with a load forecast 15 days ahead in order to improve the shutdown accuracy further.

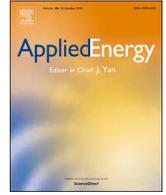
All the rules were tested and benchmarked through careful cross-validation, ensuring that the performance was never evaluated on the same data that was used to estimate parameters of the rules.

## 5.3 Main findings

Seasonal production planning is non-trivial, and a planning decision such as the summer shutdown of a large CHP plant requires careful analysis. The present study has found that the optimal dates for shutdown and start-up of a plant vary substantially from year to year. The start-up date that is optimal one year may lead to capacity problems in other years. Rare weather events that may only occur once every decade can potentially be very costly. It is therefore crucial to include as much weather data as possible in analyses of seasonal production planning decisions, to estimate the economic risk as accurately as possible.

When it comes to decision-making and estimating the savings that production planners could salvage under realistic planning condition, I have investigated decision rules with different planning horizons. Production planners, who need to plan very far ahead can salvage most of the benefit of summer shutdown by using a fixed date rule. Production planners that operate in more agile environments and are able to make decisions with shorter planning horizons can optimize the shutdown accuracy further and salvage an additional few percent of the potential savings.

The benchmarking of the decision rules showed that using only a few years of data to calibrate the rules could lead to costly surprises, again affirming the need for large amounts of weather data in seasonal production planning.



# Decision rules for economic summer-shutdown of production units in large district heating systems



Magnus Dahl<sup>a,b,\*</sup>, Adam Brun<sup>b</sup>, Gorm B. Andresen<sup>a</sup>

<sup>a</sup> Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark

<sup>b</sup> AffaldVarme Aarhus, Municipality of Aarhus, Bautavej 1, 8210 Aarhus V, Denmark

## HIGHLIGHTS

- In the study-case, the economic potential for summer-shutdown is 6.3 million €/yr.
- Optimal shutdown and start-up dates are characterized for 38 historical years.
- A fixed date rule achieves 90.7% of the potential and allows long-term planning.
- A load based rule achieves 95.8% but does not allow planning ahead.
- Using 15-day ahead forecast data increases the economic gain to 96.5% on average.

## ARTICLE INFO

### Keywords:

District heating  
Decision rules  
Long-term production planning  
Heat load forecasting

## ABSTRACT

Seasonal load variations in district heating systems are so large that some production units become superfluous during summer operations. There is great economic potential in shutting down these units during summer. The economic benefit of summer shutdown is highly dependent on the timing of the shutdown decision. The optimal shutdown and start-up dates depend on complex weather patterns and vary significantly from year to year. This study introduces three classes of decision rules to help production planners perform economically optimal summer shutdown: a fixed date rule, a heat load based rule and a load based rule augmented with weather forecasts. These decision rules are tested using 38 years of hourly weather data to simulate the heat load in Aarhus, Denmark. The large amount of weather data allows for the creation of highly robust decision rules that account for rare, but costly weather conditions. A fixed date rule allows for planning very far ahead and can reap 90.7% of the potential economic benefit of summer shutdown. A heat load based decision rule can salvage 95.8% of the potential shutdown savings at the cost of shorter planning horizons. Augmenting the load based decision rule with 15 day weather forecasts can boost the performance to 96.5%.

## 1. Introduction

The daily heat load in a district heating system varies by as much as a factor of 8 between summer and winter in a northern European climate. In continental climates, seasonal load variations are presumably even greater. In district heating systems with several production units, the seasonal load variations make some production units superfluous during summer.

Economically optimal production of heat often follows a simple merit order in which cheaper production units are given priority over more expensive ones. However, many cogeneration plants have a substantial minimum load. This means that if such a plant is kept in operation, typically to ensure low-cost security of supply, it may displace

production from a cheaper alternative. Therefore, it can be economically beneficial to completely shut down a plant during the summer period when the demand can be safely met by lower cost units. However, the challenge in performing summer shutdown of such seasonally superfluous production units lies in the timing. As an example, an early warm spring may be interrupted by a sudden cold spell, where the cost of emergency backup reserves, e.g. oil boilers, outweighs the economic benefit of an early shutdown of one of the main units. Similarly, a late start-up in fall may also require the use of costly emergency reserves if the beginning of winter is not correctly anticipated. This means that the economically optimal shutdown and start-up dates depend on complex weather patterns that vary substantially from year to year. Therefore, good decision rules are necessary to help

\* Corresponding author at: Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark.  
E-mail address: [magnus.dahl@eng.au.dk](mailto:magnus.dahl@eng.au.dk) (M. Dahl).

<http://dx.doi.org/10.1016/j.apenergy.2017.09.040>

Received 20 June 2017; Received in revised form 21 August 2017; Accepted 10 September 2017

Available online 21 September 2017

0306-2619/© 2017 Elsevier Ltd. All rights reserved.

## Nomenclature

$\sigma$	standard deviation in heat load model [ $^{\circ}\text{C}$ ]		
$\tau_1, \tau_2, \tau_3$	load thresholds for load based decision rules [MWh/h]		
$\theta_i^i$	indicator function for the availability of unit $i$		
$a$	slope in heat load model [MWh/h/ $^{\circ}\text{C}$ ]		
$c^i$	operational production cost per unit heat for production unit $i$ [€/MWh]		
$C_{\text{fall}}(t_{\text{start-up}})$	total operational production cost in the fall given $t_{\text{start-up}}$		
$[C_{\text{fall}}^{\text{no shutdown}}]$			
$C_{\text{fall}}^{\text{no shutdown}}$	total operational production cost in the fall if shutdown is not performed [€]		
$C_{\text{spring}}(t_{\text{shutdown}})$	total operational production cost in the spring given $t_{\text{shutdown}}$		
$[C_{\text{spring}}^{\text{no shutdown}}]$			
$C_{\text{spring}}^{\text{no shutdown}}$	total operational production cost in the spring if shutdown is not performed [€]		
$c_t^{\text{tot}}$	total operational heat production cost in hour $t$ [€/h]		
$h$	forecast horizon for load based decision rules with forecast		
		[h]	
	$p(T_0^i)$	probability density function for the threshold temperature in heat load model	
	$P^{\text{tot}}$	total heat load of the city [MWh/h]	
	$P_t^i$	heat production from unit $i$ in time step $t$ [MWh/h]	
	$P_0^i$	base heat load in heat load model [MWh/h]	
	$P_{\text{max}}^i$	maximum heat output from unit $i$ [MWh/h]	
	$P_{\text{min}}^i$	minimum heat output from unit $i$ [MWh/h]	
	$T_0$	threshold temperature in heat load model [ $^{\circ}\text{C}$ ]	
	$T_1, T_2, T_3$	smoothing time scales for load based decision rules [h]	
	$T_{\text{anchor}}$	anchoring time scale for load based decision rules [h]	
	$t_{\text{fixed}}$	plant shutdown/start-up time decided by the fixed date rule	
	$T_{\text{out}}$	outdoor temperature [ $^{\circ}\text{C}$ ]	
	$t_{\text{shutdown}}$	plant shutdown time	
	$t_{\text{shutdown}}^{\text{opt}}$	optimal plant shutdown time	
	$t_{\text{start-up}}$	plant start-up time	
	$t_{\text{start-up}}^{\text{opt}}$	optimal plant start-up time	

production planners perform optimal summer shutdown.

The economic potential in performing summer shutdowns of superfluous plants in large district heating systems is significant. In this paper, we use the large district heating system of Aarhus, Denmark, as a case to present decision rules that can reap almost all the benefit of shutting down a large CHP plant. This amounts to annual savings of 6.1 million €, which corresponds to about 5% of the total operational heat production cost. Environmental benefits are also significant in cases where a fossil fuel plant is shut down and the next plant in the merit order has lower carbon emissions. The economic potential is clear from the district heating perspective, but summer shutdown is economical from a system perspective as well. Combined heat and power plants (CHP) are under increasing economic pressure in many countries due to larger penetrations of renewable energy reducing electricity prices. This is especially true in summer when the heat demand is low. The overall profitability of a plant can be improved by shutting it down during the summer.

Cost-optimal operation and production planning of cogeneration systems have been studied extensively in the past. Many studies apply linear programming models (LP) [1] or mixed integer linear programming models (MILP) [2,3] to optimize the unit commitment and load dispatch in cogeneration systems. While these studies present theoretically optimal production plans, they do so under the assumption of known electricity prices and perfect load forecasts. In an actual planning situation, there will be uncertainties associated with the future heat load and electricity prices. Smart metering data are used in [4] to optimize the operation of a small district heating system with genetic optimization algorithms. Sensitivity analysis [4] and Monte Carlo simulations [5] are used to evaluate the effect of uncertain parameter estimates. Other studies account for uncertain forecasts by creating artificial price forecasts and validating the results against the assumption of perfect forecasts [6]. In [7], the authors explicitly model the uncertainties and design operational strategies based on information gap decision theory. A natural way to handle optimization under uncertainty is stochastic programming. It is used in e.g. [8] to optimize the operation of a district heating system with a strong coupling to the electricity sector through cogeneration, heat pumps and electric boilers.

None of these studies account for long production planning horizons of a month or longer. Production planning with horizons up to a month was considered in [9], and the author concluded that it is very difficult to use traditional optimization models to make relevant production plans on horizons as long as a month. In [10], a framework is introduced for modeling the planning under uncertainty on multiple time scales, the longest being more than a year. The formulation of the framework is concluded in a second paper [11], in which decision

theory is used to evaluate long-term cogeneration planning decisions.

Optimal design of cogeneration systems has been studied extensively using MILP and mixed integer nonlinear programming models. Small systems, such as a single CHP unit at a building complex, are sized in [12,13] using MILP modeling. Larger systems have been sized in [14] with a special emphasis on legal constraints. Finally, some studies include optimization of the district heating grid in the system design [15,1]. In an entirely different approach, the study in [16] evaluates optimal energy sources from the viewpoint of different stakeholders in the construction of a distributed energy system.

In this work, we use 38 years of weather data to quantify the uncertainty in a specific shutdown decision, which is the optimal summer shutdown of the large CHP plant at Studstrup, from which the city of Aarhus, Denmark, gets most of its heat. For this case, we benchmark the performance of three classes of shutdown decision rules. The first class of rules uses fixed dates, shutting down and starting up the plant on the same date every year. This class of rules has the advantage of allowing for very long planning horizons. However, it does not reap the full benefit of summer shutdown, because it is too inflexible. The second class of rules uses information about the heat load up to the point of shutdown. This class of rules performs better, but it comes at the cost of shortening the planning horizon. The third class of rules augments the load based rule with heat load forecasts up to 15 days ahead in an attempt to improve the accuracy of the shutdown even further.

The decision rules are designed to be directly applicable by system operators, as opposed to outputs of large scale MILP optimizations. MILP optimizations of the production may result in summer shutdown of certain units if start-up costs are properly accounted for. However, these optimization studies do not provide a guide to decision makers that can help with timing the decision accurately. By focusing on a single decision and creating concrete decision rules, the results of this work can be directly implemented by production planners. Although weather based modeling is not new to load modeling in district heating, see e.g. [17–21], the magnitude of our data foundation is unprecedented. With 38 years of weather data, we are able to capture effects of rare, but costly weather phenomena. Weather based modeling on this scale has previously been applied to highly renewable electricity systems to characterize transmission needs [22,23], storage [24,25] and export schemes [26].

The novelty of this work can be summarized as follows. To the best of our knowledge, we are the first to use weather based heat load modeling with such a large quantity of data. The many years of weather data allow us to capture complex patterns in the heat load to make robust decision rules. This procedure accounts for rare weather phenomena that can potentially be very costly if they are ignored. Because

of the high number of years, the analysis is able to capture rare variations in the weather that occur, e.g. once per decade, and even 100-year events have a good chance to be included. Secondly, this work focuses on the consequences of a single biannual decision and provides simple decision rules that can help production planners maximize the economic benefit by making the right decision. Our decision rules are easily implemented and rely exclusively on information that is available at the time of decision making. While based on a local case, our method and decision rules can be transferred to other locations.

The remainder of the paper is structured as follows. Section 2 describes the data foundation, the cost modeling, the heat load modeling and the decision rules. The results are presented in Section 3. Here the economic potential of summer shutdown is quantified and the decision rules are benchmarked. Finally, the paper is concluded in Section 4.

## 2. Methodology and data

In the following sections our methods and data foundation are described. First, we present the data foundation in Section 2.1. Section 2.2 details how the heat production cost is calculated with and without the unit that is shut down for summer. The modeling procedure, used to simulate the heat load based on weather data, is described in Section 2.3. Finally, in Section 2.4 we develop the shutdown decision rules and describe how they are benchmarked.

### 2.1. Heat production and weather data

The district heating system in Aarhus, Denmark has been used as a case in this study. Heat production data from three years, 2014–2016 has been used to fit and validate a heat load model. The model was fitted to measurements of the ambient air temperature from a weather station owned by the district heating company in Aarhus. The weather station is located at the heat exchanger station Brabrand Syd, and hourly measurements of the outdoor temperature in 2014–2016 were used. Hourly heat load was simulated by applying the heat load model to 38 years of weather data from NCEP CFSR [27]. The CFSR data set spans the entire globe with a spatial resolution of about 40 km and hourly time resolution. We have used data from a point in the southern part of Aarhus (N56°2'42.24", E9°59'59.95") and bias-corrected it with respect to the measurements from the Brabrand Syd weather station.

Medium-range weather forecasts for Aarhus for the spring of 2017 have been provided by the Danish Meteorological Institute (DMI). The forecasts have a horizon of 15 days and were produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [28]. The forecast data were used to create a synthetic forecast model to explore the value of augmenting decision rules with weather forecast data.

Table 1

Input parameters estimated for the Aarhus district heating production system.  $P_{min}$  and  $P_{max}$  are the minimum and maximum output of the production units when they are running.  $c$  is the operational heat production cost per unit heat. The production cost for the Studstrup CHP plant varies depending on the heat load of the plant. It is highest for low outputs and approaches the low value for high outputs.

Unit	$P_{min}$ [MWh/h]	$P_{max}$ [MWh/h]	$c$ [€/MWh]	Description
O + VE	5	10	27	Industrial waste heat from Nordalim and Jysk Miljørens
RS	5	24	15	Waste incineration in Skanderborg, Renosyd
O1 + 2	5	37	13	Waste incineration plant in Lisbjerg, unit 1 and 2
O4	5	42	13	Waste incineration plant in Lisbjerg, unit 4
BKVV	5	80	39	Straw CHP plant in Lisbjerg
BKVV BP	5	30	46	Bypass operation of BKVV
O1 + 2 BP	5	6	49	Bypass operation of O1 + 2
O4 BP	5	8	49	Bypass operation of O4
CHP	36	540	43–91	Wood pellets CHP plant in Studstrup
Other	5	515	82	Various oil boilers, electric boiler in Studstrup

### 2.2. Heat production cost

In Danish district heating systems, such as the one in Aarhus, the electricity and the heating side of a cogeneration system can be considered as two separate businesses. The heating side is non-profit by law, whereas the electricity side can be a privately owned for-profit company. The production cost is split between the electricity side and the heating side according to contracts that are typically negotiated from year to year and are based on average annual revenues from electricity sales. This cost splitting is regulated by law. Using the negotiated effective heat production cost, we can model the heat production cost in Aarhus, to a good approximation, without explicitly modeling the electricity side.

The total cost of producing heat can be divided into fixed cost and operational cost. The fixed cost is independent of the operation of the system and includes investment cost and certain maintenance cost. The operational cost, typically proportional to the produced heat, depends on fuel prices, taxes and tariff and on the cost splitting negotiated with the electricity side. The total operational production cost can be optimized by changing the system operation, whereas the total fixed cost is largely unaffected. Therefore, only the operational cost is considered in the remainder of this analysis.

Production units may have a cost associated with start-up and shutdown. When performing time coupled MILP optimization of production systems with many units, start-up and shutdown cost can be important. Including start-up and shutdown cost can help prevent instabilities in the model, where plants start up and shut down unreasonably often. If there are many small plants in the system, start-up and shutdown cost may also constitute a significant cost component. However, the composition of the Aarhus district heating system ensures that the smaller cheap units are always running and that the daily load fluctuations are accommodated by one large plant. This means that it is very rarely beneficial for units to perform start-up or shutdown. Shutdown and start-up of the large CHP plant only occur once a year in this operational scheme. Altogether, this makes the start-up and shutdown cost negligible compared to the annual savings potential of performing summer shutdown (about 6 million €).

In a large district heating system, the total operational production cost depends on which units are producing the heat, because the various production units have different production cost. Given a heat demand, we calculate total operational production cost by solving a mixed integer linear programming (MILP) problem. We let  $c^i$  be the operational production cost in €/MWh and  $P_t^i$  be the heat production in MWh/h for unit  $i$  in hour  $t$ .  $\phi_t^i$  is a binary indicator function that is 1 when the unit is running and 0 when the unit is not running.

The cost-optimal production plan for a given hour is then found by solving the problem:

$$\begin{aligned}
c_t^{\text{tot}} = \operatorname{argmin}_{P_t^i, \theta_t^i} & \sum_{i \in \text{prod. units}} c^i P_t^i \theta_t^i \\
\text{subject to} & \sum_{i \in \text{prod. units}} P_t^i \theta_t^i = P_t^{\text{tot}} \\
& P_{\min}^i \leq P_t^i \leq P_{\max}^i \quad \forall i \in \text{prod. units.}
\end{aligned} \quad (1)$$

Here,  $c_t^{\text{tot}}$  is the total operational heat production cost in €/h in hour  $t$ , and  $P_t^{\text{tot}}$  is the total heat load. The heat load is the total heat consumption plus the heat that is lost in the system. The parameters for estimating the optimal heat production cost for the Aarhus district heating production system are shown in Table 1. Bypass operation of the waste incineration plant and the straw CHP plant in Lisbjerg are treated as separate production units when determining the cost. The bypass production costs are calculated based on an average day-ahead electricity price. The production cost of the CHP plant at Studstrup varies depending on the heat load. The cost is lower at higher loads and is always below 50 €/MWh when the load is above 200 MWh/h.

The results of cost-optimal heat production can be summarized by the hourly cost plotted against the production as seen in Fig. 1. The production cost with the Studstrup CHP plant available is shown in blue, and the cost with the plant shut down is shown in green. The green cost curve has been calculated by solving (1) through a range of heat loads,  $P^{\text{tot}}$ , under the constraint that  $\theta^{\text{CHP}} = 0$ . Similarly, the blue curve has been calculated under the constraint that  $\theta^{\text{CHP}} = 1$ . It is clear that there is a crossover point when producing about 270 MWh/h at 9700 €/h. Above this heat load, it is economically favorable to have the Studstrup CHP plant available and producing. Below this load, it is cheaper to produce the heat without the Studstrup CHP plant. Thus, it may be favorable to shut down the plant entirely during a period in summer, since the heat load in Aarhus in the summer can be as low as 100 MWh/h.

Notice that the green line ends at 641 MWh/h. This is the  $N-1$  redundant capacity of the production system without the Studstrup CHP plant, i.e. the production capacity in the unfortunate case that the second largest production unit fails. The second largest production unit is the straw fired CHP plant in Lisbjerg (BKVV<sup>1</sup> + BKVV BP in Table 1).

### 2.3. Heat load model

In order to investigate the optimal summer shutdown of the CHP plant over many years, we have constructed a model for the total heat load  $P^{\text{tot}}$  as a function of the outdoor temperature. Daily average values of the heat load are shown as a function of daily average temperature in Fig. 2. The hockey stick shape of heat load plotted against temperature is well known in the district heating literature [29]. Below 15 °C, the heat load decreases linearly with the outdoor temperature. Above this point, the demand is constant to a good approximation. The behavior can be explained by the fact that below a certain temperature the heat demand is dominated by space heating, and the heat loss from buildings depends linearly on the ambient air temperature. Above a certain temperature, the heat load is dominated by hot water consumption which does not depend much on the weather. Sometimes, this relationship is modeled using a piecewise linear model [20], and sometimes using the popular degree day method [29]. We have chosen a slightly more advanced model that captures the gradual transition between these two regimes (linearly decreasing and constant). Our model, derived in the appendix, has the following functional form:

$$P^{\text{tot}}(T_{\text{out}}) = a \left[ (T_{\text{out}} - T_0) \frac{1}{2} \operatorname{erfc} \left( \frac{T_{\text{out}} - T_0}{\sqrt{2} \sigma} \right) - \frac{\sigma}{\sqrt{2\pi}} \exp \left( -\frac{(T_{\text{out}} - T_0)^2}{2\sigma^2} \right) \right] + P_0. \quad (2)$$

$T_{\text{out}}$  is the outdoor temperature.  $\operatorname{erfc}$  is the complementary error function.  $a$ ,  $T_0$ ,  $\sigma$  and  $P_0$  are parameters in the model. Estimates of the

model parameters and their physical interpretations can be found in the appendix. In the limit where  $\sigma$  goes to zero, this model reduces to a piecewise linear model with slope  $a$  for  $T_{\text{out}} < T_0$  and constant at  $P_0$  for  $T_{\text{out}} \geq T_0$ .

The red curve in Fig. 2 shows the model fitted on production data from 2014 and 2015 and temperature data from AffaldVarme Aarhus' weather station at Brabrand Syd. The model has been fitted using the feasible generalized least squares technique to account for heteroskedasticity [30].

Heat consumption varies during the day according to a characteristic daily profile [31]. This profile is characterized by a morning peak, due to many people taking hot showers in the morning around the same time. However, the intra-day heat load profile is significantly different in the weekend compared to the work days, and it also varies with the season. For the modeling in this paper, we have used 8 different profiles shown in Fig. 3. The daily average heat load, predicted using the model (2), has been converted to hourly time series by multiplication with the heat demand profiles shown in Fig. 3. The profiles were estimated from the measured heat load in 2014 and 2015. They were calculated as the mean relative deviation from the daily average production given the hour, season and whether it was a work day or weekend.

The heat load model, converting ambient air temperature from NCEP CFSR to hourly heat load, was validated on data from 2016. It yielded a mean error of 0.4 MWh/h, a root mean square error of 53 MWh/h and a mean absolute percentage error of 10.7%. The distribution of the heat load was captured well, and both mean and standard deviation of the modeled load was close to that of the measured load. It is possible to achieve higher accuracy with more elaborate time series models [21,32] or machine learning models [33,34], but for the present application simplicity was valued highly, since the focus was on long-term simulation and not short-term prediction.

Applying the heat load model to 38 years of weather data from NCEP CFSR, we created 38 years of synthetic heat load data. This is the foundation of the cost estimates in the remainder of the paper.

### 2.4. Choosing the shutdown and start-up dates

As we saw on Fig. 1, it is cheaper to produce heat without the CHP plant when the heat demand is sufficiently low. It is therefore natural to consider shutting down the plant for summer at some point during spring and starting it up again in the fall when the heat demand rises

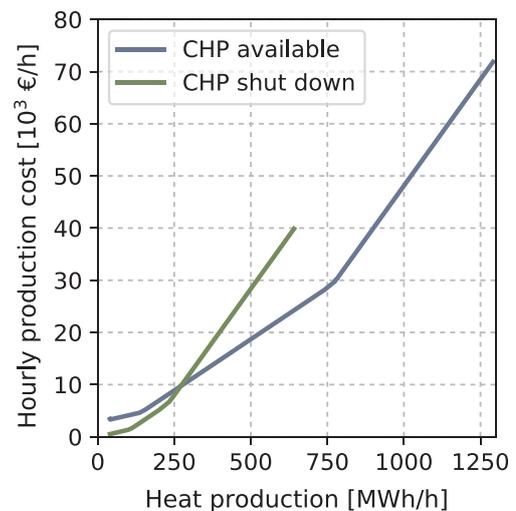


Fig. 1. The total operational heat production cost per hour as a function of the hourly heat production. The blue line shows the cost when the Studstrup CHP plant is available. The green line shows the cost when the plant is shut down. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

<sup>1</sup> From Danish: Biomassefyret KraftVarmeVærk. Literal translation: Biomass CHP plant.

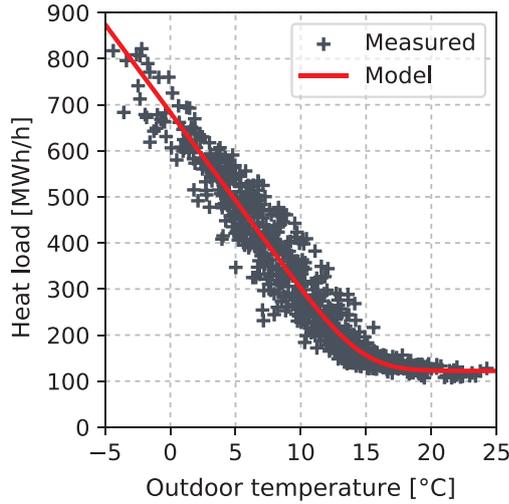


Fig. 2. Daily average heat load versus daily average ambient air temperature for 2014 and 2015 in Aarhus. The red line shows a fit of the model (Eq. (2)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

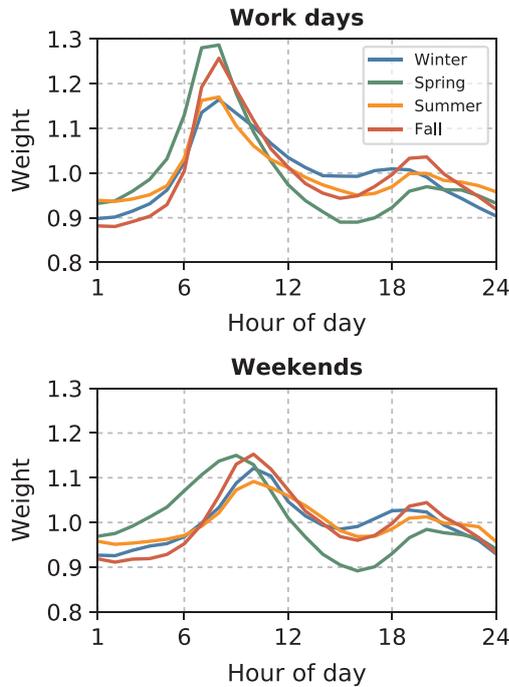


Fig. 3. Average intra-day profiles for the heat demand. The profiles are estimated from the heat demand of Aarhus in 2014 and 2015. Work days are Monday to Friday, weekends are Saturday and Sunday. For the purpose of the demand profiles the seasons are defined as winter: December to February, spring: March to May, summer: June to August and fall: September to November.

again. There are potentially large savings in performing such a shut down, but they are sensitive to the shutdown date and the start-up date. In the following analysis we consider the shutdown in the spring and the start-up in the fall separately. We define spring as the time of year before July 17th and fall as the time after. It is assumed that it is always cheapest to have the plant shut down in mid-July, and this is well justified as the heat load of Aarhus is usually well below 200 MWh/h in July.

Let  $C_{\text{spring}}(t_{\text{shutdown}})$  be the total operational cost of producing heat for Aarhus during the spring when shutting down the CHP plant at the time  $t_{\text{shutdown}}$ .  $C_{\text{spring}}(t_{\text{shutdown}})$  is measured in units of  $C_{\text{spring}}^{\text{no shutdown}}$ , the total operational production cost if the CHP plant is not shut down at

all. Completely analogously, we denote the total operational cost of producing heat in the fall by  $C_{\text{fall}}(t_{\text{start-up}})$  when starting up the CHP plant at the time  $t_{\text{start-up}}$ . The optimal shutdown and start-up times we denote  $t_{\text{shutdown}}^{\text{opt}}$  and  $t_{\text{start-up}}^{\text{opt}}$ . These are the shutdown and start-up times that minimize the production cost in the respective season. Or formally:

$$t_{\text{shutdown}}^{\text{opt}} = \underset{t_{\text{shutdown}}}{\operatorname{argmin}} C_{\text{spring}}(t_{\text{shutdown}}), \quad (3)$$

$$t_{\text{start-up}}^{\text{opt}} = \underset{t_{\text{start-up}}}{\operatorname{argmin}} C_{\text{fall}}(t_{\text{start-up}}). \quad (4)$$

This is illustrated for a representative year in Fig. 4. The gray curves show  $C_{\text{spring}}(t_{\text{shutdown}})$  and  $C_{\text{fall}}(t_{\text{start-up}})$ . The optimal dates to shut down and start up the plant are shown as yellow diamonds. In this example case, the total production cost during spring can be reduced to about 95% of the production cost of not shutting down the CHP at all. Here the full potential spring savings are about 5% of the total production cost in the reference scenario. The dark blue bar marks the date interval in which the CHP can be shut down while still realizing 90% of the potential savings. The red bar marks the interval in which it is not possible to shut down the CHP without compromising the  $N-1$  redundant capacity specification. This means that if the CHP is shut down before this date and the second largest production unit fails, it will be impossible to meet the heat demand.

#### 2.4.1. Scoring the shutdown and start-up dates

When testing the performance of decision rules for choosing the shutdown and start-up date, we used a simple scoring system. A shutdown date was scored based on how much of the potential shutdown savings were realized by shutting down the plant on a given date. Mathematically, the score is defined as:

$$\operatorname{score}(t_{\text{shutdown}}) = \frac{1 - C_{\text{spring}}(t_{\text{shutdown}})}{1 - C_{\text{spring}}(t_{\text{shutdown}}^{\text{opt}})}, \quad (5)$$

$$\operatorname{score}(t_{\text{start-up}}) = \frac{1 - C_{\text{fall}}(t_{\text{start-up}})}{1 - C_{\text{fall}}(t_{\text{start-up}}^{\text{opt}})}. \quad (6)$$

If a shutdown date scored 100%, it indicated that the full potential shutdown savings for the spring had been realized. In the Aarhus system, the potential annual savings are typically around 2.5 million € in the spring plus around 3.8 million € in the fall, if shutdown and start-up are performed on the optimal dates. In relative terms, typical savings amount to about 2% of the total annual operational cost in the spring and 3% in the fall.

#### 2.4.2. Three classes of decision rules

In this study, we benchmarked the performance of three different classes of decision rules that can help decision makers shut down and start up superfluous production units as close to the optimal dates as possible.

**2.4.2.1. Fixed date rule.** The first rule shuts down and starts up the plant on the same date every year. The fixed date was chosen as the median of the optimal dates in a training set. We have experimented with using the mean of the optimal dates, but the median performed significantly better.

**2.4.2.2. Load based rule.** From Fig. 1 we recall the crossover heat load at which it becomes cheaper to produce the heat without the CHP plant. Bearing this in mind, it is expected that more accurate shutdown and start-up can be achieved by using load based rules. Intuitively, it makes sense to shut down the plant once the total heat load of the system  $P^{\text{tot}}$ , averaged over a given period  $T$  falls below a certain threshold level  $\tau$ . Unfortunately, such a simple rule has proven to perform poorly in some years, so we have chosen a slightly more complicated class of load based rules. The idea is the same, but we demand the load averaged on

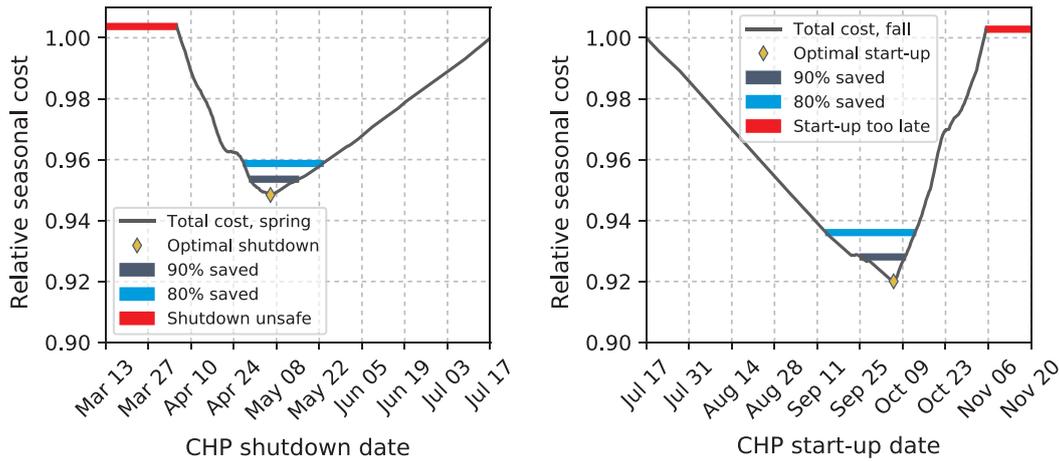


Fig. 4. Example year 2002. Total operational production cost over the season relative to the reference cost of not shutting down the CHP plant at all. The seasonal cost is shown as a function of the shutdown and start-up dates respectively. On the left:  $C_{spring}$  versus  $t_{shutdown}$ . On the right  $C_{fall}$  versus  $t_{start-up}$ . The optimum is indicated with a diamond. The light blue interval marks the ranges of dates within which at least 80% of the potential savings can be realized. The dark blue interval shows the interval for at least 90% savings. The red interval indicates the dates for which the CHP has to be running, in order to maintain the  $N-1$  redundant capacity. This is one example year, similar figures for all 38 years are summarized in Fig. 5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

two different time scales  $T_1$  and  $T_2$  must cross the thresholds  $\tau_1$  and  $\tau_2$ . Formally, the rule shuts down the plant in the spring the first time step  $t$  in which all of these three conditions are met:

$$\begin{aligned} \frac{1}{T_1} \sum_{t'=t-T_1}^t P_{t'}^{tot} &\leq \tau_1, \\ \frac{1}{T_2} \sum_{t'=t-T_2}^t P_{t'}^{tot} &\leq \tau_2, \\ t &\geq t_{fixed} - T_{anchor}. \end{aligned} \quad (7)$$

The last condition anchors the shutdown date to the date chosen by the fixed date rule. This is a safety precaution that ensures that the load based rule does not shut down the plant more than  $T_{anchor}$  periods (typically 2 weeks), before the fixed date rule would shut down the plant. The rule is not very sensitive to the choice of  $T_{anchor}$  as long as it is not too small. The averaging time spans:  $T_1$  and  $T_2$  are tuning parameters of the rule and are fixed by the user. The thresholds  $\tau_1$  and  $\tau_2$  are determined by training the rule on a set of years. In the training, the thresholds are determined by minimizing the sum of squared distances to the optimal shutdown times. Averaging the heat load over two different time scales in (7) makes the rule more robust. Fluctuations in the heat load occur both on a diurnal time scale due to consumer behavior and a synoptic time scale due to changes in the weather. The double averaging of the heat load allows the rule to react to changes on different time scales and makes it robust to fluctuations on just one time scale.

Completely analogously, the load based rule for deciding to start up the plant in the fall is defined as follows. In the fall (after July 17th), the plant is started up again in the first time step  $t$  that fulfils all of these three conditions:

$$\begin{aligned} \frac{1}{T_1} \sum_{t'=t-T_1}^t P_{t'}^{tot} &\geq \tau_1, \\ \frac{1}{T_2} \sum_{t'=t-T_2}^t P_{t'}^{tot} &\geq \tau_2, \\ t &\geq t_{fixed} - T_{anchor}. \end{aligned} \quad (8)$$

Again,  $T_1$ ,  $T_2$  and  $T_{anchor}$  are tuning parameters, and the thresholds  $\tau_1$  and  $\tau_2$  are determined through least squares training.

The best tuning parameters have been found through a grid search and 7-fold cross-validation on the 35 years 1979–2013. For the spring shutdown rule we found  $T_1 = 2$  h and  $T_2 = 504$  h and for the fall start-up

rule we found  $T_1 = 2$  h and  $T_2 = 3$  h. Notice that there is a great difference in the optimal parameters to use. This can be attributed to the fact that the rule only relies on information about the heat load prior to the decision time. Therefore, the spring and fall decisions are fundamentally different in nature.

2.4.2.3. Load based rule with 15 day forecast. The load based decision rule described above relies exclusively on the heat load prior to the time of the decision. Using weather forecasts to predict the heat load up to 15 days ahead may improve this class of decision rules. A load based decision rule augmented with weather forecasts 15 days ahead can be formulated as follows. The plant is shut down the first time step  $t$  when the load based conditions (7) are true and an additional forecast based condition is true:

$$\frac{1}{T_3} \sum_{t'=t+h-T_3}^{t+h} P_{t'}^{tot} \leq \tau_3. \quad (9)$$

Here,  $h$  is the forecast horizon of 360 h (15 days). Similarly, in the fall, the plant is started up when all the start-up conditions in (8) are met and the forecast based condition

$$\frac{1}{T_3} \sum_{t'=t+h-T_3}^{t+h} P_{t'}^{tot} \geq \tau_3 \quad (10)$$

is met. The averaging periods  $T_1$ ,  $T_2$  and  $T_3$  are chosen through tuning, and the best parameters have been found to be  $T_3 = 336$  h in the spring and  $T_3 = 24$  h in the fall.  $T_1$  and  $T_2$  were chosen as before. The thresholds  $\tau_1$ ,  $\tau_2$  and  $\tau_3$  were again determined through least squares training by minimizing the squared distance to the optimal shutdown/start-up date.

### 2.4.3. Simulated weather forecasts

Since actual weather forecasts for 38 years have not been available to us, we have simulated weather forecasts by adding artificial forecast errors to the temperature data. Based on 15 day ensemble forecasts of the ambient temperature in Aarhus we have built a model for the daily forecast error. The error at forecast horizons from 1 to 15 days can be approximated well by a 15 dimensional multivariate normal distribution. This modeling procedure preserves the variance of the forecast error on different horizons and captures the covariance structure of the error by design. Simulated forecast errors were added to the temperature data, and used when training and testing the decision rules augmented with forecast data.

### 3. Results and discussion

Inspecting Fig. 4, it is clear that for this example year the cost-optimal decision would be to shut down the CHP plant on May 6 and start it up again on October 6. However, shutting down the plant in the 16 day window between April 29 and May 15 would still reap 90% of the potential benefit of the shutdown in the spring. If we are satisfied with realizing 80% of the benefit of the shutdown, the possible shutdown dates constitute an interval of almost 4 weeks. However, caution must be observed. By shutting down before April 6 (red interval), we

would be gambling with the security of supply. In this interval, there is not enough production capacity if the second largest production unit fails.

#### 3.1. Realizing the potential of summer shutdown

The example year shown in Fig. 4 is modeled on weather data from 2002. A similar curve for the cost versus shutdown/start-up date has been calculated for all the 38 years in the data set. While the overall shape of the curves is similar from year to year, both the location and

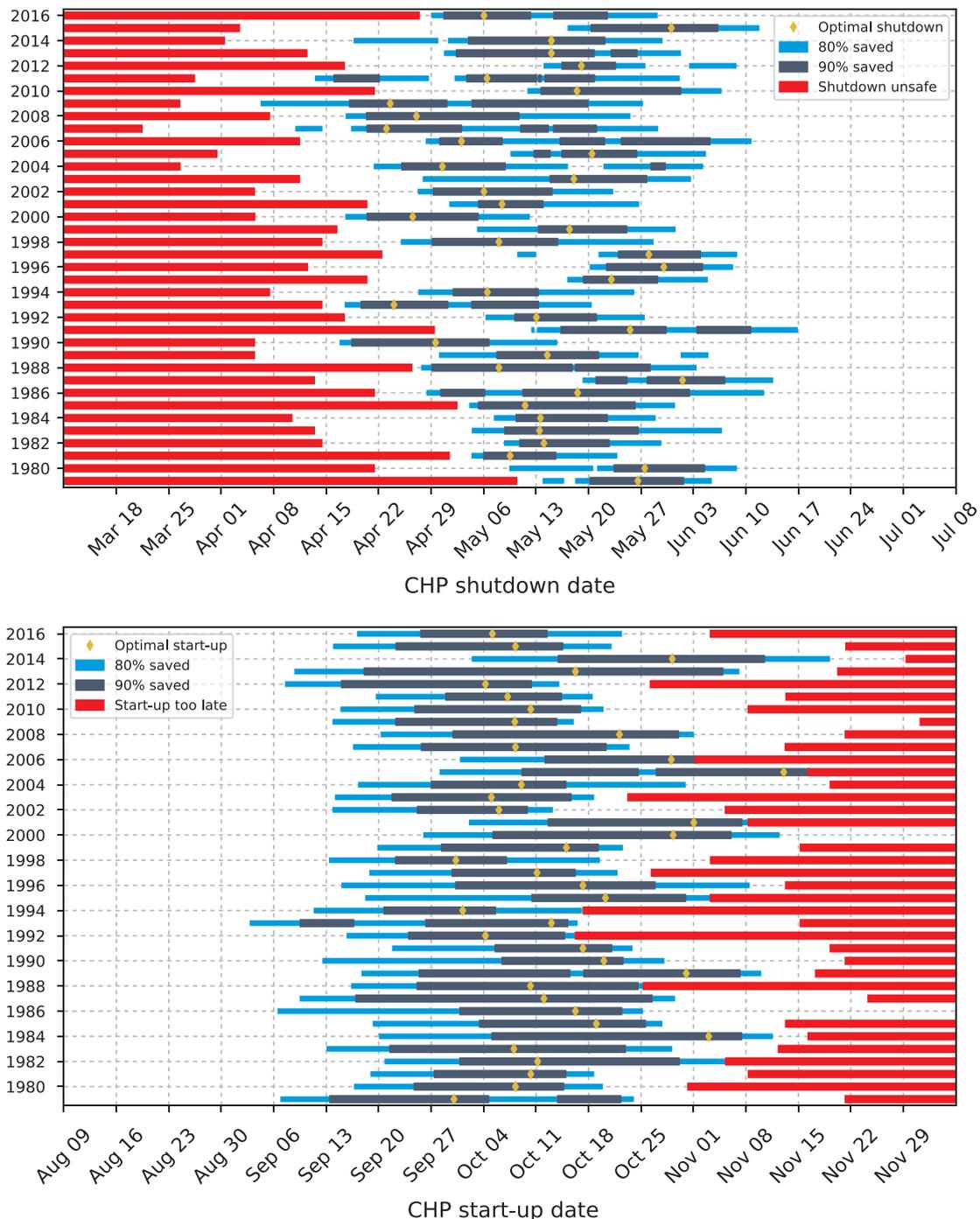


Fig. 5. Optimal shutdown and start-up dates for all years. On the top: Optimal dates to shut down the CHP plant in the spring are shown as yellow diamonds. On the bottom: Optimal dates to restart the plant in the fall. The light and dark blue bars indicate the time intervals within which a shutdown/start-up can realize 80% or 90% of the total economic benefit of shutting down the plant. The red bars indicate periods in which the plant must be running, to ensure  $N-1$  redundant capacity in the system. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

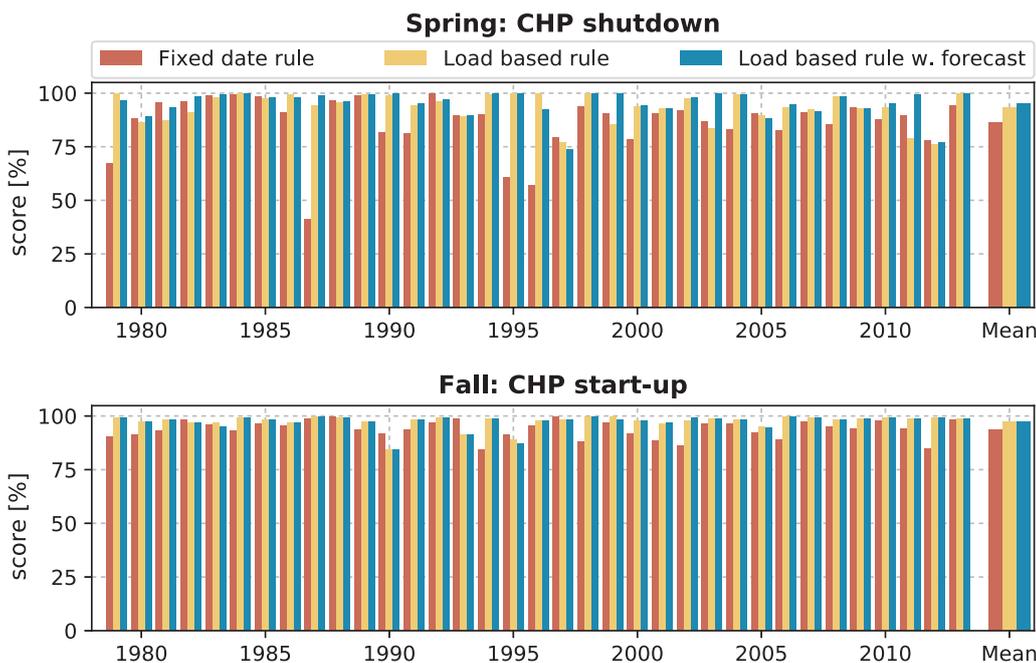


Fig. 6. Scoring of the three decision rules in the cross-validation on the years 1979–2013. A score of 100% indicates that the decision rule was able to realize the full economic potential of performing summer shutdown in the respective season. On the very right a summary of all the scores is shown. Scores for the blind test years 2014, 2015 and 2016 are shown in Table 2.

the depth of the minimum vary substantially. The depth of the minimum indicates the potential savings from summer shutdown. The variation in the location of the minimum reflects the challenge of shutting down and starting up the plant at the optimal times.

In the spring, the average savings that can be achieved from optimal shutdown is 3.5% of the total spring production cost without shutdown. It varies between 1.6% and 5.3%. In the fall, the average savings amount to 8.6% and varies between 5.2% and 15.4%. The apparent difference between spring and fall is due to the fact that both of the highest load months, January and February, are situated in the spring period. In absolute terms, this means that there is an annual savings potential of 6.3 million € on average for the case of Aarhus. However, in some years it may be as low as 4.5 million € or as high as 7.6 million €.

Cost curves similar to Fig. 4, for all the 38 years, have been summarized in Fig. 5. The optimal shutdown and start-up dates are indicated for each year. Intervals in which a shutdown or start-up would reap 80% or 90% of the potential savings are shown as light and dark blue bars. For each year, the red bar indicates the time of year in which a shutdown of the CHP would be unsafe from a capacity perspective.

The optimal shutdown date in the spring varies significantly from year to year. The median shutdown date is on May 13, but varies between April 23 and June 1. In the fall, the median start-up date is October 11, but varies in the range September 30 to November 13. While the optimal shutdown and start-up times and vary over a range of about 6 weeks, there is an overlap of the 80% intervals for most of the years. This is especially true for the start-up in the fall. It is therefore possible to select a fixed start-up date in the fall which, for all years, reaps at least 80% of the potential shutdown savings over the fall period. Shutdown in the spring is slightly more difficult, but there are still fixed shutdown dates that reap at least 80% of the benefit in most years. Great care must be taken when choosing a fixed shutdown or start-up date, since dates that are optimal in some years fall within the unsafe (red) period in other years.

### 3.2. Fixed date shutdown/start-up

A good choice of date, for the fixed date decision rule is the median of the optimal dates in the training set. We score the performance of the decision rules based on the percentage of the potential summer shutdown savings that are realized when using the rule. In the process of

selecting and validating the decision rules, we have used 7-fold cross-validation on the 35 years 1979–2013. The years 2014–2016 were withheld for a final blind test. In Fig. 6, the red bars show the validation scores using the fixed date rule. On average, the fixed date rule was able to realize 86.3% of the potential savings in the spring and 93.8% of the potential savings in the fall. The rule performs well for most years, but there are a few outliers. Most notable is the shutdown in the spring of 1987, in which less than half of the potential shutdown savings were realized. This is the weakness of the fixed date rule. The performance is usually good, but there is a risk of foregoing large saving in some years. The fact that deciding the shutdown in the spring is more difficult than deciding the start-up in the fall was anticipated from Fig. 5.

The performance of the fixed date rule is decent, and it has the advantage of allowing for planning very far ahead. This is a major advantage in large production systems where the planning of shutting down plants for maintenance must often be scheduled and coordinated far in advance. While the fixed date rule performs reasonably well, there is room for improvement, and salvaging a few more percentage points can be justified when the total potential savings are 6.3 million € per year.

### 3.3. Load based shutdown/start-up

In the hopes of improving on the performance of the fixed date rule, we have introduced a class of load based rules. This class of rules, described in Section 2.4.2, uses information about the heat load up to the point of the decision to shut down or start up the plant. The yellow<sup>2</sup> bars in Fig. 6 show the performance of the load based rule in the cross-validation. The average performance is improved to 93.4% in the spring compared to 86.3% from using the fixed date rule. In the fall, we see an improvement to 97.4% from 93.8% with the fixed date rule. In certain years, such as 1987, 1995 and 1996, we see very drastic improvements of 40 percentage points or more in the spring. All in all, there is significant economic potential in switching from the fixed date rule to a load based rule. However, it comes at the cost of decreasing the planning horizon, since the load based decision rule uses data up to the point of the shutdown/start-up decision.

<sup>2</sup> For interpretation of color in Fig. 6, the reader is referred to the web version of this article.

### 3.4. Load based shutdown/start-up with forecast

Despite the load based decision rule outperforming the fixed date rule by a decent margin, there is still a small room for improvement. An improvement of 1 percentage point corresponds to more than 60,000 € in annual savings on average, so there is an incentive to pursue the last few points up to 100%. Therefore, we augment the load based decision rule with weather forecasts up to 15 days ahead. Utilizing forecasts to supplement the decision with information about the future has the potential to improve the performance of the rules. However, there is a risk that the forecast uncertainty is so large that the decision making does not improve significantly.

The red bars in Fig. 6 show the scores of the load based rule augmented with weather forecasts. Across the whole cross-validation, the forecast augmented rule performs very similarly to the load based rule. There are a few years in which the spring shutdown is improved by more than 15 percentage points e.g. 1999, 2003 and 2011. Augmenting the load based rule results in an average improvement of 1.9 percentage points to 95.3% in the spring and a small drop in performance from 97.4% to 97.3% in the fall. Overall, there is a small performance improvement to gain by augmenting the load based forecast rule with weather forecast information.

When benchmarking the three decision rules, we applied 7-fold cross-validation to the years 1979–2013. Three test years, 2014–2016, were withheld during the specification and tuning of the rules. The results of training the rules on the cross-validation period and testing them on the test period are shown in Table 2 along with a summary of the cross-validation. The test set was made quite small, so as to utilize the data as well as possible for the tuning and cross-validation. This has the downside that the test performance is relatively volatile. The cross-validation results on the other hand are very stable. The performance on the test years show that although a rule performs well overall, there may be outliers. The fixed date rule performed extremely well in 2014 while the load based rule was less impressive. However, we saw the same general trend in the test and training scores as we saw in the cross-validation. It is harder to achieve good performance in the spring shutdown than in the fall start-up.

### 3.5. Rare weather events

Using as many as 38 years of weather data has the benefit that it is easier to catch rare weather events in the modeling. In this context, rare weather events cause the optimal start-up or shutdown date to be unusually early or unusually late. A good example is the optimal start-

**Table 2**

Summary benchmarking of the three decision rules. A score of 100% indicates that the full savings potential has been realized for the respective season. The blind test was performed by training the rules on the years 1979–2013 and testing on 2014–2016. The validation score comes from 7-fold cross-validation on the years 1979–2013. In parentheses we indicate 1 standard deviation on the mean values for test, training and cross-validation.

Decision rule type	Spring shutdown score [%]			Fall start-up score [%]		
	Fixed date	Load based	Load based w. forecast	Fixed date	Load based	Load based w. forecast
Test year 2014	99.0	80.5	97.6	88.0	98.8	98.8
Test year 2015	71.6	71.1	94.3	97.1	95.1	95.1
Test year 2016	88.3	88.5	91.9	93.0	96.1	96.1
Mean test score ( $1\sigma$ )	86.3 (11.2)	80.1 (7.1)	94.6 (2.4)	92.7 (3.7)	96.7 (1.7)	96.7 (1.6)
Mean training score ( $1\sigma$ )	86.9 (12.0)	93.8 (6.6)	96.2 (4.3)	94.4 (4.4)	97.4 (3.2)	97.1 (3.4)
Mean validation score ( $1\sigma$ )	86.3 (12.6)	93.4 (6.7)	95.3 (6.0)	93.8 (4.1)	97.4 (3.1)	97.3 (3.3)

up date in the fall of 2005. The bottom of Fig. 5 shows that the optimal start-up date fell on November 13 in 2005. This is an unusually late start-up date. The date is an outlier by the Tukey method [35], and it is more than 2 standard deviations above the mean of the optimal dates. The start-up date fell late in 2005 due an unusually warm October followed by a sudden drop in the outdoor temperature in the middle of November. Despite the uncommon weather pattern in the fall of 2005, the start-up rules performed decently in the cross-validation. The fixed date rule achieved a score of 92.1% and the load based rule could salvage 94.9% of the potential fall savings.

Another benefit of using many years of weather data is that the robustness of the decision rules can be gauged more realistically. Inspecting the top part of Fig. 6, we can imagine a decision maker in 1986 using data from 1980–1986 to decide on a shutdown rule. This decision maker, might have chosen to use the fixed date rule to shutdown the plant in spring, since its performance seemed comparable to the load based rules. The following year, 1987 the fixed date rule performed poorly and the decision maker did not foresee this, because only data from the early 1980s had been used. Events where the fixed date shutdown rule has scored below 75% have occurred only four times in the 38 year period, and only the year 1987 has performed below 50%. These events are so rare that there is a large chance to miss them if only 5 or 10 years of data are used when designing the decision rules.

## 4. Conclusion

We have shown how production cost of district heating can be significantly reduced by performing summer shutdown of superfluous production units. This kind of shutdown can be performed in any district heating system with multiple production units. Using Aarhus, Denmark as a case, we have shown that the economic potential of shutting down a large CHP plant to be around 6.3 million € each year. This amounts to about 5% of the annual operational production cost.

The economic benefit of the summer shutdown is highly dependent on the timing of the shutdown and start-up of the plant. 38 years of weather data has been used to build robust decision rules to help production planners shut down the plant in the most optimal way.

A production planner who needs to plan very far ahead can benefit from using a fixed date rule. The fixed date rule shuts down and starts up the plant on the same dates every year. This rule is capable of reaping most of the potential savings of summer shutdown (90.7%). Production planners operating in a more agile setting with shorter planning horizons can benefit from using one of the load based decision rules. The summer shutdown savings can be increased by using these rules (95.8%), but it comes at the cost of short planning horizons. Finally, it was shown that load based decision rules can be improved a little by augmenting them with mid-range weather forecast. This decision rule is capable of reaping 96.5% of the total annual benefit of performing summer shutdown.

The decision rules presented in this work either have infinite planning horizons (fixed date rule) or no planning horizon (load based rules). There is a trade-off between the planning horizon and the performance of the rules. In the current paper, we have focused on improving the performance of the rule by using forecasts. Instead, medium-range weather forecasts can be used to extend the planning horizon. Continued research into the topic should explore this trade-off.

When using such a large amount of weather data, shifts in the Earth's climate may begin to have an effect. Future studies should attempt to quantify how climate change will affect operational decisions in district heating systems.

Seasonal heat storage could function as a supplement to summer shutdown of seasonally superfluous production units. Weather dependent storage operation is crucial in district heating systems with large solar heating production capacity and seasonal storage. The methodology presented in this paper can be applied to create operational

decision rules for optimal operation of seasonal heat storage facilities using large amount of weather data.

The decision rules presented here can be adopted by district heating operators and trained on measured load data. If several years of load data are not available, a simple weather based heat load model can be used to train the rules.

## Acknowledgements

This study is part of the READY project (Resource Efficient cities

## Appendix A

The model for the heat load of a city as a function of the outdoor temperature in Eq. (2) is derived in this appendix.

The hockey stick shape of the heat load versus temperature (see Fig. 2) is well known in district heating literature [29]. Below a threshold temperature,  $T'_0$ , the heat load,  $P$ , decreases linearly with the temperature,  $T_{out}$ . In this regime, the heat load is dominated by space heating, and heat losses from buildings are linear with the outdoor temperature to a good approximation. Above the threshold temperature, the heat load is constant,  $P_0$ , mainly due to hot water consumption. For a single building, a piecewise linear function is a good approximation to the heat load:

$$P_{pw.lin.}(T_{out}; T'_0) = \begin{cases} a(T_{out} - T'_0) + P_0 & \text{for } T_{out} < T'_0 \\ P_0 & \text{for } T_{out} \geq T'_0 \end{cases} \quad (11)$$

Here  $a$  is the slope of the linear part of the function. The threshold temperature,  $T'_0$ , varies from building to building depending on specific installations and the behavior of the residents. However, in a city with a large number of buildings, we can assume that it is normally distributed according to the central limit theorem. If the threshold temperature,  $T'_0$ , is normally distributed with mean value  $T_0$  and variance  $\sigma^2$  its probability density function,  $p$ , reads:

$$p(T'_0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(T'_0 - T_0)^2}{2\sigma^2}\right). \quad (12)$$

In order to find a functional form for the heat load of an entire city,  $P^{tot}$ , we integrate the heat load of the individual buildings with the probability density function:

$$\begin{aligned} P^{tot}(T_{out}) &= \int_{-\infty}^{\infty} P_{pw.lin.}(T_{out}; T'_0) p(T'_0) dT'_0 \\ &= a \left[ (T_{out} - T_0) \frac{1}{2} \operatorname{erfc}\left(\frac{T_{out} - T_0}{\sqrt{2}\sigma}\right) \right. \\ &\quad \left. - \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(T_{out} - T_0)^2}{2\sigma^2}\right) \right] + P_0. \end{aligned} \quad (14)$$

Here,  $\operatorname{erfc}$  is the complementary error function. This model recreates the hockey stick shape that can be seen in heat load data aggregated on the city level e.g. in Fig. 2. In the limit  $\sigma \rightarrow 0$ , where all heat consumers are alike, the model reduces to (11). Furthermore, (14) behaves asymptotically like (11) when  $T_{out} \ll T_0$  and when  $T_{out} \gg T_0$ .

The parameters of the model have been fitted on daily heat load and daily average temperatures for Aarhus in 2014 and 2015 using the technique of feasible generalized least squares. The parameters were estimated to:

$$a = -38.3 \text{ MWh/h/}^\circ\text{C}, \quad T_0 = 14.6 \text{ }^\circ\text{C}, \quad P_0 = 122 \text{ MWh/h} \quad \text{and} \quad \sigma = 2.67 \text{ }^\circ\text{C}.$$

## References

- [1] Aringhieri Roberto, Malucelli Federico. Optimal operations management and network planning of a district heating system with a combined heat and power plant. *Ann Operat Res* 2003;120(1–4):73–199.
- [2] Dvořák Michal, Havel Petr. Combined heat and power production planning under liberalized market conditions. *Appl Therm Eng* 2012;43:163–73.
- [3] Thorin Eva, Brand Heike, Weber Christoph. Long-term optimization of cogeneration systems in a competitive market environment. *Appl Energy* 2005;81(2):152–69.
- [4] Fang Tingting, Lahdelma Risto. Genetic optimization of multi-plant heat production in district heating networks. *Appl Energy* 2015;159:610–9.
- [5] Koltsaklis Nikolaos E, Kopanos Georgios M, Georgiadis Michael C. Design and operational planning of energy networks based on combined heat and power units. *Indust Eng Chem Res* 2014;53(44):16905–23.
- [6] Fang Tingting, Lahdelma Risto. Optimization of combined heat and power production with heat storage based on sliding time window method. *Appl Energy* 2016;162:723–32.
- [7] Aghaei Jamshid, Agelidis Vassilios G, Charwand Mansour, Raeisi Fatima, Ahmadi Abdollah, Nezhad Ali Esmaeel, et al. Optimal robust unit commitment of CHP plants in electricity markets using information gap decision theory. *IEEE Trans Smart Grid* 2016.
- [8] Nielsen Maria Grønnegaard, Morales Juan Miguel, Zugno Marco, Pedersen Thomas Engberg, Madsen Henrik. Economic valuation of heat pumps and electric boilers in the danish energy system. *Appl Energy* 2016;167:189–200.
- [9] Dotzauer Erik. Experiences in mid-term planning of district heating systems. *Energy* 2003;28(15):1545–55.
- [10] Carpaneto Enrico, Chicco Gianfranco, Mancarella Pierluigi, Russo Angela. Cogeneration planning under uncertainty: part I: Multiple time frame approach. *Appl Energy* 2011;88(4):1059–67.
- [11] Carpaneto Enrico, Chicco Gianfranco, Mancarella Pierluigi, Russo Angela. Cogeneration planning under uncertainty. Part II: Decision theory-based assessment of planning alternatives. *Appl Energy* 2011;88(4):1075–83.
- [12] Beihong Zhang, Weiding Long. An optimal sizing method for cogeneration plants. *Energy Build* 2006;38(3):189–95.
- [13] Gamou Satoshi, Yokoyama Ryohei, Ito Koichi. Optimal unit sizing of cogeneration systems in consideration of uncertain energy demands as continuous random variables. *Energy Convers Manage* 2002;43(9):1349–61.
- [14] Lozano Miguel A, Ramos Jose C, Serra Luis M. Cost optimization of the design of CHCP (combined heat, cooling and power) systems under legal constraints. *Energy* 2010;35(2):794–805.
- [15] Chinese Damiana. Optimal size and layout planning for district heating and cooling networks with distributed generation options. *Int J Energy Sector Manage* 2008;2(3):385–419.
- [16] Ghafghazi Saeed, Sowlati Taraneh, Sokhansanj Shahabaddine, Melin Staffan. A multicriteria approach to evaluate district heating system options. *Appl Energy* 2010;87(4):1134–40.
- [17] Werner Sven. The heat load in district heating systems, PhD thesis. Chalmers tekniska högskola; 1984.
- [18] Madsen Henrik, Sejling Ken, Søgaard Henning T, Palsson Olafur P. On flow and

- supply temperature control in district heating systems. *Heat Recov Syst CHP* 1994;14(6):613–20.
- [19] Heller AJ. Heat-load modelling for large systems. *Appl Energy* 2002;72(1):371–87.
- [20] Dotzauer Erik. Simple model for prediction of loads in district-heating systems. *Appl Energy* 2002;73(3–4):277–84.
- [21] Nielsen Henrik Aalborg, Madsen Henrik. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy Build* 2006;38(1):63–71.
- [22] Rodríguez Rolando A, Becker Sarah, Andresen Gorm B, Heide Dominik, Greiner Martin. Transmission needs across a fully renewable European power system. *Renew Energy* 2014;63:467–76.
- [23] Becker Sarah, Rodríguez Rolando A, Andresen Gorm B, Schramm Stefan, Greiner Martin. Transmission grid extensions during the build-up of a fully renewable pan-European electricity supply. *Energy* 2014;64:404–18.
- [24] Rasmussen Morten G, Andresen Gorm B, Greiner Martin. Storage and balancing synergies in a fully or highly renewable pan-European power system. *Energy Policy* 2012;51:642–51.
- [25] Pleßmann Guido, Erdmann Matthias, Hlusiak Markus, Breyer Christian. Global energy storage demand for a 100% renewable electricity supply. *Energy Proc* 2014;46:22–31.
- [26] Rodríguez Rolando A, Dahl Magnus, Becker Sarah, Greiner Martin. Localized vs. synchronized exports across a highly renewable pan-European transmission network. *Energy Sustain Soc* 2015;5(21).
- [27] Saha S, Moorthi S, Pan H-L, Wu X, Wang J, Nadiga S, et al. The NCEP climate forecast system reanalysis. *Bull Am Meteor Soc* 2010;91(8):1015–57.
- [28] Persson Anders. User guide to ECMWF forecast products. ECMWF; 2015.
- [29] Frederiksen Svend, Werner Sven. District heating and cooling. Studentlitteratur; 2013.
- [30] Wooldridge Jeffrey M. *Introductory econometrics: a modern approach*. Nelson Education; 2015.
- [31] Gadd Henrik, Werner Sven. Daily heat load variations in Swedish district heating systems. *Appl Energy* 2013;106:47–55.
- [32] Dahl Magnus, Brun Adam, Andresen Gorm B. Using ensemble weather predictions in district heating operation and load forecasting. *Appl Energy* 2017;193:455–65.
- [33] Kato Kosuke, Sakawa Masatoshi, Ishimaru Keiichi, Ushiro Satoshi, Shibano Toshihiro. Heat load prediction through recurrent neural network in district heating and cooling systems. In: 2008 IEEE international conference on systems, man and cybernetics. IEEE; 2008. p. 1401–6.
- [34] Idowu Samuel, Saguna Saguna, Åhlund Christer, Schelén Olov. Forecasting heat load for smart district heating systems: a machine learning approach. In: 2014 IEEE international conference on smart grid communications (SmartGridComm); 2014.
- [35] Tukey John W. *Exploratory data analysis*. Mass.: Reading; 1977.

## Chapter 6

# Production system planning

### 6.1 Motivation

The three previous chapters have been focused on district heating production systems and their optimal operation. Operational uncertainty has been quantified with respect to short-term and seasonal production planning as well as system operation. In this chapter, I move on from the current heat production system and begin to quantify the uncertainty related to planning of future district heating systems. The topic of my last research paper is the cost-optimal capacity allocation and technology choices for district heating production systems.

Investments in energy production capacity carry significant risk. Changing energy demands and changing prices on electricity, fuels and technologies all affect the economy of heat production units. Given increasing efforts to reduce carbon emissions globally, energy system planners need to account for possible bans on fossil fuels. In the following article, I investigate how sensitive a cost-optimal heat production system is to changing investment cost, fuel cost and electricity prices in a transition towards fossil free heat production.

Another effect of the global push to reduce carbon emissions is a rapid increase in wind power generation in the North European electricity systems (Jones et al., 2018). Current pricing structures result in very low, even negative, electricity prices when the market is periodically flooded with cheap wind power. As we move towards higher wind penetrations in the electricity system, this kind of price volatility is likely to increase (Woo et al., 2011). In addition to the sensitivity to electricity pricing levels, I therefore examine the impact of wind power dominated electricity pricing on the optimal production system.

Summing up, my last research article characterizes how cost-optimal heat production systems change when transitioning away from fossil fuels, under different electricity pricing schemes. Furthermore, it is quantified how sensitive the cost-optimal system is to changes in investment cost, fuel cost and electricity prices.

### 6.2 Methods

Since planning of district heating systems often takes place on an urban level, the modeling was focused on the local city-wide heat production system. District

heating production systems are typically coupled to a larger electricity system, so electricity prices and demands for electricity and heat were treated as external boundary conditions imposed on the district heating system.

### Capacity optimization

The optimal capacity allocation in the Aarhus district heating system was found by solving a large linear optimization problem. The total annual investment and operational costs were minimized with respect to system operation and installation of production and storage capacities. The study included only well-established technologies, and a range of heat only boilers, CHP units, power-to-heat technologies and heat storages were available for installation in the system.

The capacity allocation and system operation was posed as a linear programming problem, optimizing an entire year's operation with hourly time resolution. The heat load was sourced from the Aarhus district heating system, and electricity prices and demand from the local transmission system operator. A large number of constraints ensured physically correct behavior of the CHP units and heat storages and the model was validated by reproducing the production system operation of Aarhus in the year 2015.

### Electricity pricing schemes

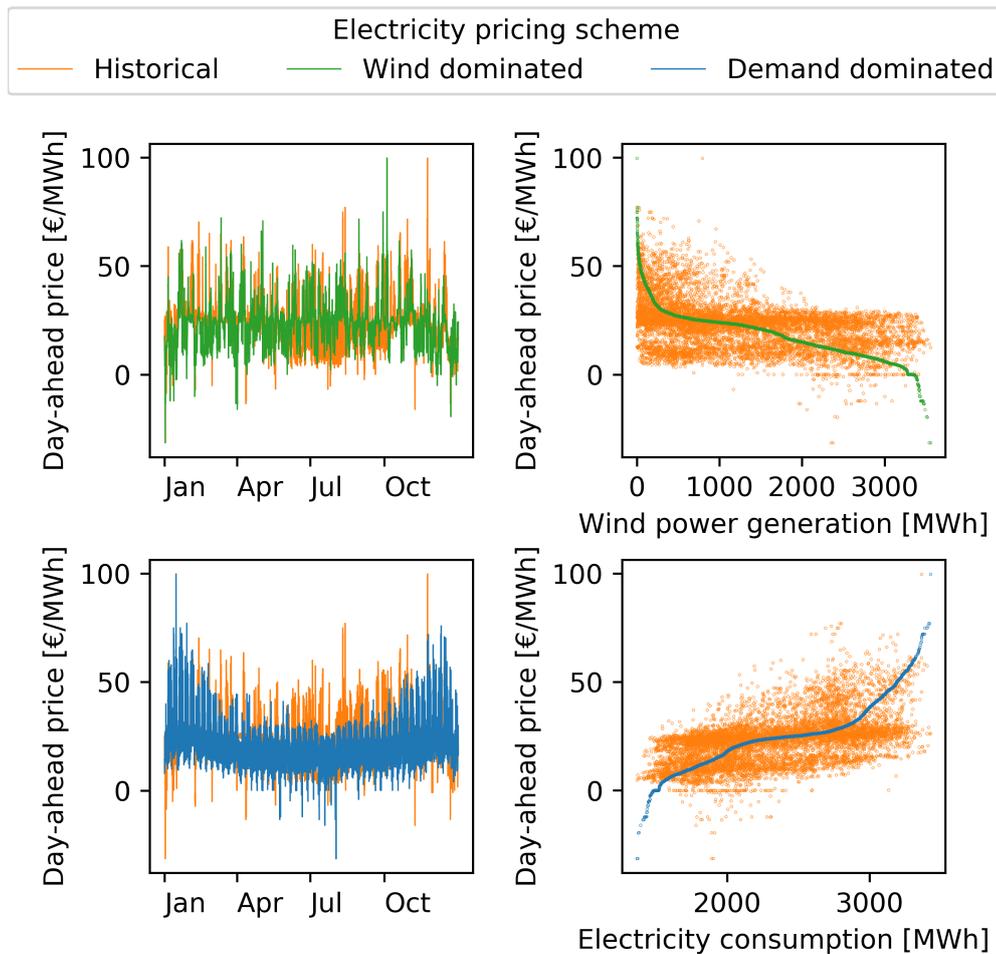
Three different electricity pricing schemes and their effects on the optimal heat production system were studied in the paper: Historical, wind dominated and demand dominated electricity pricing. The historical pricing scheme was identical to the electricity prices in the West Danish day-ahead market DK1 in 2015. Figure 6.1 shows the price in the historical pricing scheme in orange.

A possible future, in which wind power generation has a greater influence on electricity prices, was investigated in a wind dominated pricing scheme, plotted in green. In this pricing scheme the hour with the highest wind power generation was assigned the lowest electricity price. The top right scatter plot in Figure 6.1 illustrates this procedure. By sorting the historical electricity price in order to make the negative correlation between price and wind power generation as strong as possible, I preserved the full distribution of the electricity price over the year. This facilitated a fair comparison between the different electricity pricing schemes.

The same methodology was applied to create a demand dominated pricing scheme, plotted in blue. In this pricing scheme, the electricity price correlated strongly with the electricity demand. This was intended to simulate a future with larger amounts of dispatchable power generation compared to variable renewable energy sources such as wind or solar power generation.

### Sensitivity analysis

In optimization studies of energy systems, sensitivity analyses are often limited or completely omitted (Lindenberger et al., 2000; Münster et al., 2012; Ortiga et al., 2013; Buoro et al., 2014). This may be due to high computational costs



**Figure 6.1:** Three different electricity pricing schemes with hourly electricity prices based on the West Danish day-ahead market DK1 in 2015.

of running the optimizations, but it is problematic since sensitivity studies provide valuable information about the robustness of the results with respect to changing model parameters. Decision makers in charge of planning future energy production systems are faced with large uncertainties on fundamental cost assumptions which propagate into uncertainty on the optimal planning decision.

In the present study, these uncertainties were quantified through an extensive multivariate sensitivity analysis. It was investigated how variations in investment cost, fuel cost and electricity prices affected the optimal capacity allocation in a large district heating system.

One-at-a-time sensitivity analyses, in which one input parameter is varied while the others are fixed, can be overly optimistic because they do not adequately explore the input space and fail to account for interactions between inputs (Czitrom, 1999). I therefore employ a more sophisticated type of multivariate sensitivity analysis based on the experimental design Latin hyper cube sampling. This type of sampling varies parameters simultaneously while ensuring that the

samples are representative of the full multivariate variability (McKay et al., 1979).

The sensitivity analysis in this study consisted of 200 different sensitivity scenarios, including many different combinations of rising and falling electricity prices and fuel and investment costs for individual technologies.

### 6.3 Main findings

The cost-optimizations in this study did not account for local taxes or regulations, but the results can serve as a guiding point for energy planners and lawmakers. Going forward, a city-wide district heating system coupled to a regional or national electricity system can benefit from installing a substantial amount of large-scale heat pumps supplemented by large heat storages. If fossil fuels are allowed in the future, coal fired CHP seems to play a significant role alongside heat pumps in cost-optimal heat production systems. However, the exact capacity mix between coal CHP and heat pumps is highly sensitive to changes in investment cost, fuel prices and electricity prices.

If fossil free heat production is desired, the total heat demand can be met by heat pumps, if the storage capacity is approximately doubled. Going fossil free does increase the total system cost slightly, but it significantly reduces the uncertainties in the capacity allocation. The uncertainty on the total system cost, however, is increased when going fossil free, as the cost of heat becomes more dependent on the electricity prices.

The optimal technology choices are highly stable under changing cost assumptions: coal CHP, heat pumps and heat storage pits in a fossil fuels scenario; heat pumps and pit storages in a fossil free scenario.

A wind dominated electricity pricing scheme will increase the optimal heat pump capacity and reduce the total heating system cost. This is because this pricing scheme generally has lower electricity prices in the winter when the heat load is high. A demand dominated pricing scheme will also increase the need for heat pumps, but it may increase the system price.

# Cost sensitivity of optimal sector-coupled district heating production systems

Magnus Dahl<sup>a,b</sup>, Adam Brun<sup>b</sup>, Gorm B. Andresen<sup>a</sup>

<sup>a</sup>*Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark*

<sup>b</sup>*AffaldVarme Aarhus, Municipality of Aarhus, Bautavej 1, 8210 Aarhus V, Denmark*

---

## Abstract

Goals to reduce carbon emissions and changing electricity prices due to increasing penetrations of wind power generation affect the planning and operation of district heating production systems. Through extensive multivariate sensitivity analysis, this study estimates the robustness of future cost-optimal heat production systems under changing electricity prices, fuel cost and investment cost. Optimal production capacities are installed choosing from a range of well-established production and storage technologies including boilers, combined heat and power (CHP) units, power-to-heat technologies and heat storages. The optimal heat production system is characterized in three different electricity pricing scenarios: Historical, wind power dominated and demand dominated. Coal CHP, large heat pumps and heat storages dominate the optimal system if fossil fuels are allowed. Heat pumps and storages take over if fossil fuels are excluded. The capacity allocation between CHP and heat pumps is highly dependent on cost assumptions in the fossil fuel scenario, but the optimal capacities become much more robust if fossil fuels are not included. System cost becomes less robust in a fossil free scenario. If the electricity pricing is dominated by wind power generation or by the electricity demand, heat pumps become more favorable compared to cogeneration units. The need for heat storage more than doubles, if fossil fuels are not included, as the heating system becomes more closely coupled to the electricity system.

*Keywords:* District heating, Energy production, Optimization, Cost sensitivity, Fossil free

---

## 1. Introduction

District heating systems are facing a new reality on multiple fronts. Ambitious global efforts to decrease carbon emissions call for the transformation of heat production systems away from fossil fuels and towards fossil free alternatives. In modern district heating systems combined heat and power (CHP) plants form the backbone of the production system and usually provide a majority of the heat. Coal and gas fuelled CHP is cheaper than biomass based CHP but problematic from a carbon emissions perspective. At the same time, electricity systems are quickly adopting large amounts of wind power generation, which reduces the economic feasibility of CHP generation by periodically lowering electricity prices [1, 2]. Power-to-heat technologies benefit from this development, especially in combination with heat storage technologies.

In this study, we explore how the cost-optimal compositions of a city-wide heat production system changes when moving into a fossil-free future. The effect of electricity pricing dominated by wind generation or by electricity demand is investigated, and the results are corroborated by extensive sensitivity analysis. We use the district heating system of Aarhus, Denmark as a study case, providing the heat load and the validation scenario.

District energy systems are often planned and operated on a city level. Therefore, it makes sense to model the district heating production system coupled to a larger electricity system. Taking the city's point of view in the

modeling allows us to give recommendations for energy planners under different external conditions, such as the state of the regional or national electricity system.

In [3], Lund et al. compared two different approaches to energy system modeling: simulation and optimization. Simulation studies simulate and envisage the behavior of the system under a set of operating conditions defined by the user. Scenario based modeling, e.g. in EnergyPLAN, is an example of simulation studies. In optimization studies, the values of a number of decision variables are computed to minimize a certain objective function subject to constraints. A common example is allocation of production capacities in order to minimize system cost. Both modeling paradigms have their merits, and in this study we combine the two in orders to find cost-optimal system configurations in different scenarios. These scenarios include: allowing fossil fuels, excluding fossil fuels, historical electricity pricing, wind dominated electricity pricing and demand dominated electricity pricing. Combining the two approaches, we provide recommendations that are relevant to decision makers under different planning conditions. We indicate the robustness of the recommendations under changing cost assumptions by means of thorough sensitivity analysis.

Capacity optimization studies in district energy systems are plentiful in the literature. Our system optimization includes well-established technologies such as different boilers, CHP units, electric boilers, heat pumps and heat

storages. Operations and capacities of CHP units have been optimized in [4, 5] using fossil fuels and in [6, 7] using biomass. The economic feasibility of large heat pumps for district heating systems have been investigated carefully in [8], taking day-to-day operational uncertainty into account through stochastic programming. The benefit of long-term heat storage in district heating systems has been studied in [9] and heat storage tanks have been compared to using the building mass for heat storage in [10].

Energy systems, in which it is important to model system nonlinearities, possibly making the objective function non-convex can be optimized using global optimization approaches such as genetic algorithms [7, 11, 12]. However, these approaches can be slow and run the risk of not finding the global minimum. In [9] the capacity and operation of CHP plants are optimized as mixed integer nonlinear programming (MINLP) problems, and the authors highlight some of the potential pitfalls of non-convex optimization. In cases where the energy system behavior can be reasonably linearized, the optimization speed can be decreased. Not surprisingly, mixed integer linear programming (MILP) [13, 14] and linear programming (LP) [15, 16] models are widespread in production capacity optimization and operational optimization. A thorough review of optimization studies in trigeneration systems (electricity, heating, cooling) can be found in [17].

In this study, we pose capacity and operational optimization as an LP problem and validate the resulting system operation against actual operational data for the city of Aarhus using a methodology similar to [18]. Even large LP problems with hundreds of thousands of variables and millions of constraints can be solved deterministically in relatively short time, assuming they are feasible and bounded. This property allows us to perform extensive sensitivity analysis of the cost assumptions of the model. In many optimization studies, model runs are very computationally expensive, which can severely limit the feasibility of large sensitivity analyses. In [16] the sensitivity analysis is limited to varying the fuel prices and CO<sub>2</sub> prices up and down by 50%. Most studies that do include sensitivity analysis of the model assumptions, only vary the input parameters one at the time [7, 11, 13]. One-at-a-time sensitivity analysis has the disadvantage, that it only explores a very small part of the possible input space and fails to account for interactions between input parameters [19]. In this study, we perform an extensive multivariate sensitivity analysis including 200 points. These points are sampled using an experimental design called Latin hypercube sampling (LHS) [20] in order to better capture to full variability of the cost-parameters of the model and thoroughly test the robustness of the results with respect to changes in electricity prices, fuel cost and investment cost.

A number of studies explore the effects of changing electricity prices on the economy of CHP units [12] or entire district energy systems [8, 18]. As in [13], we model the district heating system as a price-taker, that does not affect the electricity prices. We employ a novel way of con-

structing electricity price scenarios based on historical day-ahead prices, that preserves the distribution of the prices, but changes the autocorrelations. This methodology allows us to construct wind dominated electricity prices or demand dominated electricity prices, and facilitates fair comparison between these scenarios.

Some capacity optimization studies include local regulatory constraints [7, 21]. Regulations, tariffs and taxations are left out of our modeling, except for a possible ban on fossil fuels. In this way, our results represent taxation-neutral economically optimal energy systems, which can serve as a guiding point for energy planners and lawmakers.

Summing up, we demonstrate the robustness of economically optimal heat production systems under changing cost assumptions in the transition away from fossil fuels. In addition, the effects of changing influences in the electricity market are explored using a new methodology which allows for fair comparison between pricing scenarios, because it preserves the electricity price distribution.

The rest of the paper is structured as follows. In Section 2 the system model is described and validated, and the electricity pricing scenarios and sensitivity analysis are outlined. The results are presented in Section 3, and the paper is concluded in Section 4. Finally, in the Appendix the full mathematical formulation of the model can be found.

## 2. Methodology

This section describes how we have modeled the energy production system of a city with district heating coupled to a larger electricity transmission area. In Figure 1, the conceptual overview of the modeled energy system is sketched. The focus in this work is the optimal capacity configuration of a such a city, with regards to CHP production, heat only boilers, power-to-heat technologies and heat storages. The system operation and production capacity installation is co-optimized as described in Section 2.1. In Section 2.1.1 the optimization model is validated against the actual energy system operation of the city of Aarhus in 2015 and in Section 2.2 the various production and storage technologies in the capacity optimization are described. The implementation of the electricity market and three different electricity pricing schemes is covered in Section 2.3. Finally, Section 2.4 describes an extensive cost sensitivity analysis that qualifies the robustness of the final results.

### 2.1. Production system optimization

The modeling in this study is based on the city of Aarhus in Denmark. Aarhus is a city of about 340,000 people and almost all buildings in the city are heated through an extensive district heating system. The basis of this model is the hourly heat load of Aarhus from 2015.

The core of the work is a linear programming (LP) optimization problem, in which the heat and electricity

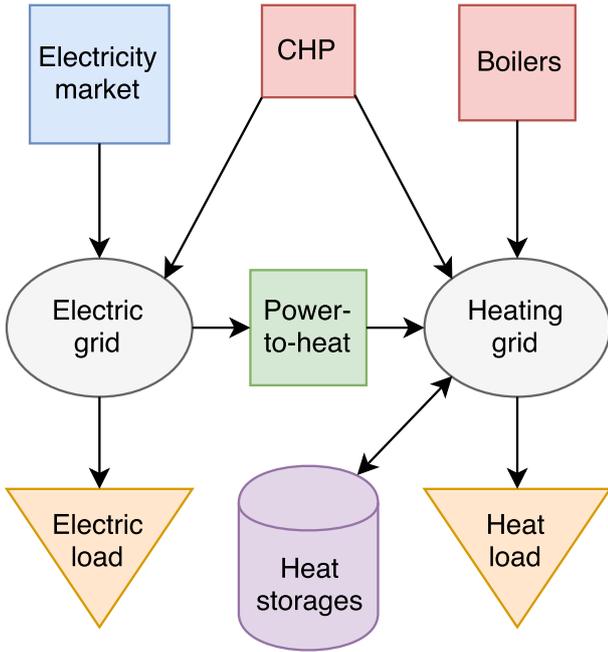


Figure 1: Conceptual overview of the energy system model. Arrows indicate directed energy flows.

production are co-optimized with heat storage operation and the installation of different types of heat and electricity production capacity. The total production system cost is minimized, including investment cost, operational and maintenance cost and fuel cost on all production units.

The full optimization problem, is formulated in the appendix, where the objective function (1) is minimized under the constraints (2-14). The system cost is minimized under a number of constraints. Energy balance constraints ensure that the total heat demand of the city is met in each hour and that the total electrical load is met as well. A number of capacity constraints ensure that all production units operate within their capacity. Finally, cogeneration production units are imposed with further constraints, depending on whether they are extraction-condensing plants or back-pressure plants with bypass.

The formulation of the capacity optimization as an LP problem means that well-established very fast techniques can be used to solve the problem in relatively short time. A scenario can be optimized in less than 10 minutes on a regular laptop as of 2018, which allows for extensive sensitivity analysis of the problem. The whole model has been implemented in Python for Power System Analysis (PyPSA) [22] and solved with the commercial Gurobi Optimizer [23].

We optimize the operation of the production and storage units on an hourly timescale throughout a full year. By using a full year, the system is operated through a representative range of the heat and electricity load. This is especially important for the heat load, as it varies by more than a factor of 8 over the course of a year [24].

### 2.1.1. Model validation

In order to validate the operational part of the optimization model, we reconstruct the operation of the Aarhus heat production system in 2015. In this operational optimization, we lock the capacities of each unit to the actual production capacity of the system. We cannot disclose the full technical specification of the Aarhus production plants, but we can summarize the most important aspects. In 2015, the base heat load was provided by 94.5 MW of waste incineration CHP, with the capability to boost the heat production to 112 MW by bypassing the turbines. A wood chips boiler of 24 MW supplemented the waste incineration as base load. A large coal CHP plant was in charge of maintaining the load balance in the system, and had a heat production capacity of 968 MW and an electrical production capacity of 707 MW. A number of peak-load oil boilers, with a total capacity of 435 MW were available in case of extremely cold weather or fallouts of other production units. Finally, a heat storage tank capable of storing 2,000 MWh was available.

When comparing the optimization results to the actual operation of the system, it is important to bear in mind that the optimization is based on the 2015 day-ahead electricity prices and operates with perfect foresight. This potentially makes the storage operation more optimal than what can be achieved by actual system operators.

On Figure 2, we see a comparison of the duration curves of the production and storage units in the Aarhus system of 2015. There is a good correspondence between the simulated and realized duration curves both in shape and magnitude. There are some smaller discrepancies, most notably in the operation of the oil boilers and the storage operation. The oil boilers are not used at all in the simulation, but in reality some oil was used. This is because the CHP plant fell out for a small period during 2015, and the excess load had to be covered with oil boilers. The difference in the shape of the storage operation, is due to tighter constraints on the storage heat uptake and dispatch in the simulation, compared to reality. It was necessary to use tighter constraints in the optimization to compensate for the perfect foresight.

A summary of the total annual heat output for each of the units can be found in Table 1. The waste incineration delivered 6% more heat in the simulation than in reality. The waste incineration is the cheapest unit and the simulation does not account for revision periods in which a unit is taken out for repairs and maintenance. This explains the excess production from waste in the simulation compared to reality. Likewise, the wood chips boiler also delivered more heat in the optimization compared to reality. This is because the wood chips boiler in reality was shut down during the summer in order to avoid competing with the waste incineration. The summed storage uptake and dispatch in the simulation are very similar to reality. The storage operation in reality was not cyclical, which explains the apparently positive net heat output from the

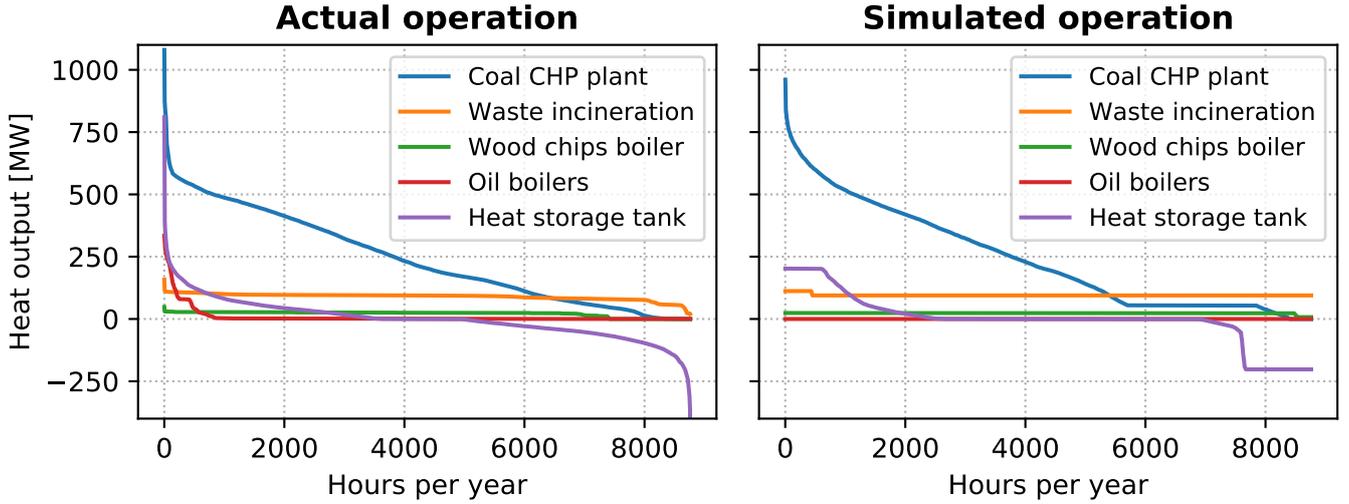


Figure 2: Duration curves for the heat output of the main production and storage units in the Aarhus heat production system of 2015. On the left, the actual operation of the units is shown; on the right are the results of the operational optimization.

storages.

Table 1: Comparison between the actual and simulated operation the Aarhus heat production system as of 2015. For each major production unit, the total annual heat production is shown. The heat storage dispatch is shown with a positive sign and storage uptake is shown with a negative sign.

	Total heat output in 2015	
	Actual [GWh]	Simulated [GWh]
Coal CHP plant	$2.11 \times 10^3$	$2.12 \times 10^3$
Waste incineration	785	836
Wood chips boiler	183	206
Oil boilers	72.5	0.00
Heat storage dispatch	241	241
Heat storage uptake	-240	-246

We cannot disclose the total electricity production in Aarhus, because it would be easy to reconstruct which plants delivered how much electricity. But, we have made the comparison between the actual and simulated electricity production and there is a difference of less than 1% [25].

All in all, we cannot expect perfect correspondence between the actual operation and the optimized operation, since the actual operation is decided by production planners without perfect knowledge of the system, and because revision periods and accidental fallouts are not included in the simulation. However, the correspondence is good enough that the operation of production units and storages can be considered realistic, also when we move on to capacity optimizations.

## 2.2. Heat production and storage technologies

The technologies we have chosen to include in the capacity optimization are all well-established technologies,

that have been implemented in district heating systems before. The production units can be divided into three types: Heat only boilers, CHP plants and power-to-heat technologies. Geothermal production technologies have not been included since their feasibility and cost are highly location dependent. Solar thermal technologies have not been included due to their high cost and negative synergies with waste incineration [26]. The financial and technical data about the heat production technologies included in the capacity optimization are summarized in Table 2.

Heat only boilers (Table 2a) are generally cheap and produce heat for the heating system by burning fuel. From an exergetic perspective, boilers are not ideal as they convert high exergy fuel into low exergy heat. We implement three common boiler types in the capacity optimization: Wood chips boilers, gas boilers and oil boilers. While wood chips boilers are preferable due to lower carbon emissions, the investment cost is significantly higher compared to gas and oil boilers. We have omitted wood pellet and straw boilers, as they are generally not economically competitive on a large scale due to the higher fuel cost [26].

As heat production units, CHP plants (Table 2b) are much more expensive than boilers, but they have the advantage that they deliver both heating and electricity. This is beneficial from an exergy perspective. Even in a future with large amounts of cheap wind power, the electricity system is likely to need dispatchable backup power, e.g. from CHP plants [30]. Our capacity optimization implements six different CHP technologies. Coal and wood pellets plants are both extraction-condensing CHP plants fired by pulverized fuel. The modeling includes three different gas fired plants: gas engines, simple cycle turbines and combined cycle turbines. Finally, a straw fired back-pressure CHP plant has been implemented, inspired by the newly opened (2017) straw fired plant in the Aarhus

Table 2: Financial and technical data for the energy production and storage technologies in the simulation based on [26, 27]. Fuel costs are sourced from [28] and missing parameters have been estimated based on the Aarhus heat production system. Efficiencies are given in terms of the lower calorific value of the fuel. All financial data for CHP technologies are in terms of electricity production and capacity. CapEx denotes the investment cost per unit capacity, OpEx denotes operation and maintenance cost and are split into a fixed annual part and a variable part depending on the operation.  $\eta^{\text{boiler}}$  is the efficiency of the boilers, and  $\eta^{\text{el}}$  is the electrical efficiency of the CHP plants.  $\alpha$  denotes the power-to-heat ratio of CHP plants in back-pressure operation and  $\zeta$  is the specific electrical power loss [29]. For power-to-heat technologies,  $\eta^{\text{boiler}}$  or COP is the ratio of produced heat to consumed electricity. For the heat storage units,  $\eta^{\text{stand}}$  is the fraction of the energy content that is lost through standing heat losses in each time step.

(a) Boiler technologies (fuel based).

Boiler type	Fuel cost [€/MWh <sub>fuel</sub> ]	CapEx [M€/MW <sub>heat</sub> ]	OpEx <sub>fixed</sub> [k€/MW <sub>heat</sub> /yr]	OpEx <sub>variable</sub> [€/MWh <sub>heat</sub> ]	Lifetime [yr]	$\eta^{\text{boiler}}$ [-]
Wood chips	24	0.8	0	5.4	20	1.08
Gas	20	0.06	2	1.1	25	1.03
Oil	46	0.06	2	0.26	25	0.94

(b) CHP technologies.

CHP type	Fuel cost [€/MWh <sub>fuel</sub> ]	CapEx [M€/MW <sub>el</sub> ]	OpEx <sub>fixed</sub> [k€/MW <sub>el</sub> /yr]	OpEx <sub>variable</sub> [€/MWh <sub>el</sub> ]	Lifetime [yr]	$\eta^{\text{el}}$ [-]	$\zeta$ [-]	$\alpha$ [-]
Straw	21	4.0	40	6.4	25	0.29	0.15	0.48
Wood pellets	25	2.0	57	2.0	40	0.46	0.15	0.75
Gas (simple cycle)	19	0.60	20	4.5	25	0.39	0.15	0.95
Gas (combined cycle)	19	0.90	30	4.5	25	0.55	0.15	1.7
Gas engines	19	1.0	10	5.4	25	0.44	0.15	0.9
Coal	9.2	1.9	32	3.0	40	0.46	0.15	0.75

(c) Power-to-heat technologies.

Power-to-heat type	CapEx [M€/MW <sub>heat</sub> ]	OpEx <sub>fixed</sub> [k€/MW <sub>heat</sub> /yr]	OpEx <sub>variable</sub> [€/MWh <sub>heat</sub> ]	Lifetime [yr]	$\eta^{\text{boiler}}/\text{COP}$ [-]
Electric boilers	0.07	1.1	0.5	20	0.98
Compression heat pumps	0.7	2.0	2	25	3.5

(d) Heat storage technologies.

Heat storage type	CapEx [€/m <sup>3</sup> ]	Storage capacity [MWh <sub>heat</sub> /m <sup>3</sup> ]	Lifetime [yr]	$\eta^{\text{stand}}$ [-]
Storage tanks	210	0.07	20	$1.4 \times 10^{-3}$
Storage pits	35	0.07	20	$1.4 \times 10^{-3}$

district heating system.

Power-to-heat technologies (Table 2c) have increasing potential in district heating applications as electricity markets are periodically flooded with large amounts of cheap wind power [8]. The capacity optimization includes simple electric boilers and compression heat pumps. Large-scale heat pumps for district heating have already been implemented in district heating systems (e.g. in Stockholm) [31], and we assume a low-temperature heat source such as seawater and a conservative coefficient of performance (COP) of 3.5. Depending on the temperature of the heat source and the district heating supply temperature, the COP can be significantly higher [27].

Finally, two different heat storage technologies (Table 2d) have been implemented: heat storage tanks and seasonal pit storages. Heat storage tanks are already common in many district heating systems around the world, including the one in Aarhus. Pit storages are gaining ground and two examples of large heat storage pits are located in Marstal ( $75 \times 10^3 \text{ m}^3$ ) and Dronninglund ( $60 \times 10^3 \text{ m}^3$ ) [32] in Denmark, and they are significantly cheaper than storage tanks for large storage volumes.

Besides delivering heat for room heating and hot water consumptions, the district heating system of Aarhus and many other large district heating systems serve another crucial societal function. Municipal waste is incinerated and the excess heat is used for district heating and electricity generation. The waste incineration needs of a city like Aarhus are unlikely to change significantly in the coming years, so the actual 2015 waste incineration capacity was therefore included in all the simulations with fixed capacity and optimized operation.

### 2.3. Electricity pricing schemes

In the modeling, the West Danish electricity market (DK1) has been implemented with historical day-ahead prices for 2015. The market is implemented in the model as a simple generator with practically unlimited production capacity, capable of delivering electricity at the spot price in the relevant time step. Effectively, this lets local CHP units deliver electricity when their production price is below the spot price and it allows power-to-heat technologies to use electricity at market price. See Figure 1.

Besides the 2015 electricity pricing (*Historical electricity pricing*), we have constructed two artificial pricing scenarios. In a market with abundant amounts of electricity, the supply side will dominate. In the current and future North European energy system, the market will periodically be dominated by large amounts of wind power [33]. Prices will go down when wind power is abundant, and when the wind settles down, prices go up. We call this scenario the *Wind dominated electricity pricing*. In the other extreme, if the energy system is not dominated by variable renewable energy sources, and the electricity demand is primarily covered by dispatchable generation with a less volatile price, the electricity price will be dominated by the demand. In hours with high demand, prices go up

and vice versa. We call this scenario the *Demand dominated electricity pricing*.

In the capacity optimization, we implement the wind and demand dominated pricing scenarios based on the historical 2015 electricity price, wind power production and electricity demand. All the data has been sourced from the local transmission system operator (TSO) Energinet [34]. We employ a methodology that preserves all moments of the distribution of the electricity price time series. In this way, there is the same mean, variance, skewness etc., but the autocorrelation is lost. A wind dominated price time series is obtained by sorting the original price such that the highest price is relocated to the hour with the lowest wind power production, the second highest price is relocated to the second lowest wind hour and so on. This process preserves the total value in the electricity market and facilitates fair comparison between the different pricing scenarios. In the historical pricing scenario, the Pearson correlation coefficient between the day-ahead electricity price and the wind power production was  $-0.40$ . In the wind dominated scenario it is  $-0.91$ .

The same methodology is used to create a demand dominated pricing scheme. This time, the electricity prices are relocated such that the highest price falls in the hour with highest demand. The correlation between demand and price goes from  $0.57$  in the historical scenario to  $0.95$  in the demand dominated scenario.

### 2.4. Sensitivity analysis

Large scale modeling studies are haunted by the fact that they require many different input parameters to define the model. These parameters may be difficult to accurately estimate or may be subject to change. It is therefore crucial to perform sensitivity analysis to investigate how robust the final results are to changes in input parameters. It is a well known problem in numerical modeling that hypervolumes grow exponentially with the dimensionality of input spaces. This is known as the curse of dimensionality, and it means that exhaustive searches through input spaces very quickly become infeasible as the number of input parameters grow, especially if model evaluations are time-costly. One technique to deal with this problem is Latin hypercube sampling (LHS), which is a random multivariate sampling method that ensures that the samples are representative of the real variability of the variables [20].

In this study, the sensitivity analysis is focused on the cost assumptions. We have run the system capacity optimization with 200 different perturbations of the CapEx and fuel cost assumptions shown in Table 2 as well as the mean electricity price<sup>1</sup>. The 200 points were generated using Latin hyper cube sampling to ensure a representative

---

<sup>1</sup>In the sensitivity analysis, the entire electricity price time series was scaled up or down with the same factor, drawn from the Latin hypercube sample.

sample of the input space. The LHS points were transformed via the inverse cumulative distribution to be normally distributed around the initial value (Table 2), with a standard deviation of 10% of the initial value. This leaves about 5% chance of the value being perturbed by more than 20%.

This methodology ensures that we investigate a representative sample of the many combinations of changes in cost assumptions. It is very important to be thorough in this kind of analysis as a rise in coal prices combined with a drop in electricity prices may yield very different results from a rise in both or from a drop in the cost of e.g. biomass.

We do not perform sensitivity analysis on the technical parameters, as most technical changes can be reduced to equivalent changes in cost assumptions.

### 3. Results

In this section we present the cost-optimal heat production capacities for the case of Aarhus embedded in the West Danish electricity market DK1. In the base scenario, all the production and storage technologies from Table 2 have been included. This case is compared to a fossil free scenario, in which all fossil fuel technologies are excluded from the capacity optimization.

Figure 3a shows optimal heat production and storage capacities for the base scenario, and the fossil free scenario is shown below in Figure 3b. The orange bars depict the optimal capacities obtained using the historical day-ahead electricity prices from 2015. This is the historical electricity pricing scheme. A first look at Figure 3 reveals that only a few technologies are assigned nonzero capacities in the cost optimization. The waste incineration is not a part of the capacity optimization, but its operation is optimized. Waste incineration aside, the heat demand is covered by a combination of coal fired CHP and compression heat pumps in the fossil fuel scenario. In the fossil free scenario, the entire heat demand is covered by heat pumps supplemented by waste incineration. Pit heat storage is the only storage technology that is utilized, which is not surprising given its low costs. Heat storage tanks may still have a place in district heating production systems, especially in places where pit storages are infeasible due to space, temperature or pumping requirements. There are two main differences between the production system configuration in the fossil fuel scenario and the fossil free scenario. The first is that the coal CHP capacity is replaced, almost one to one with heat production capacity from heat pumps. The second is that the optimal amount of heat storage doubles.

The total annual cost of the production system of the city includes investment cost, operation and maintenance cost, fuel cost and the cost of electricity for heat pumps, minus the value of generated electricity. We compare the total annual system cost to a reference system, consisting of the actual installed production capacities in Aarhus in

2015 (see Section 2.1.1), using optimized operation, the historical electricity pricing and the financial and technical data from Table 2. The difference in annual system cost between the reference and the capacity optimized system is shown in Table 3. Negative values indicate that the capacity optimized system is cheaper.

In the first row of Table 3 we see that the capacity optimized fossil fuel system reduces the cost of the production system by 8.4 million € every year, a cost reduction of about 12%. It is possible to construct a fossil free production system that is cheaper than the reference, in this case only 5.3 million € per year cheaper. The cost difference of about 3 million € between the fossil fuel and fossil free scenario can be attributed to the very low fuel price of coal.

It is important to note that the capacity optimized scenarios do not include redundancy or extra capacity for exceptionally cold years. The cost comparison in Table 3 should therefore not be interpreted a *savings potential* by transforming the energy system, but rather serve as a consistent cost comparison between the capacity optimized scenarios.

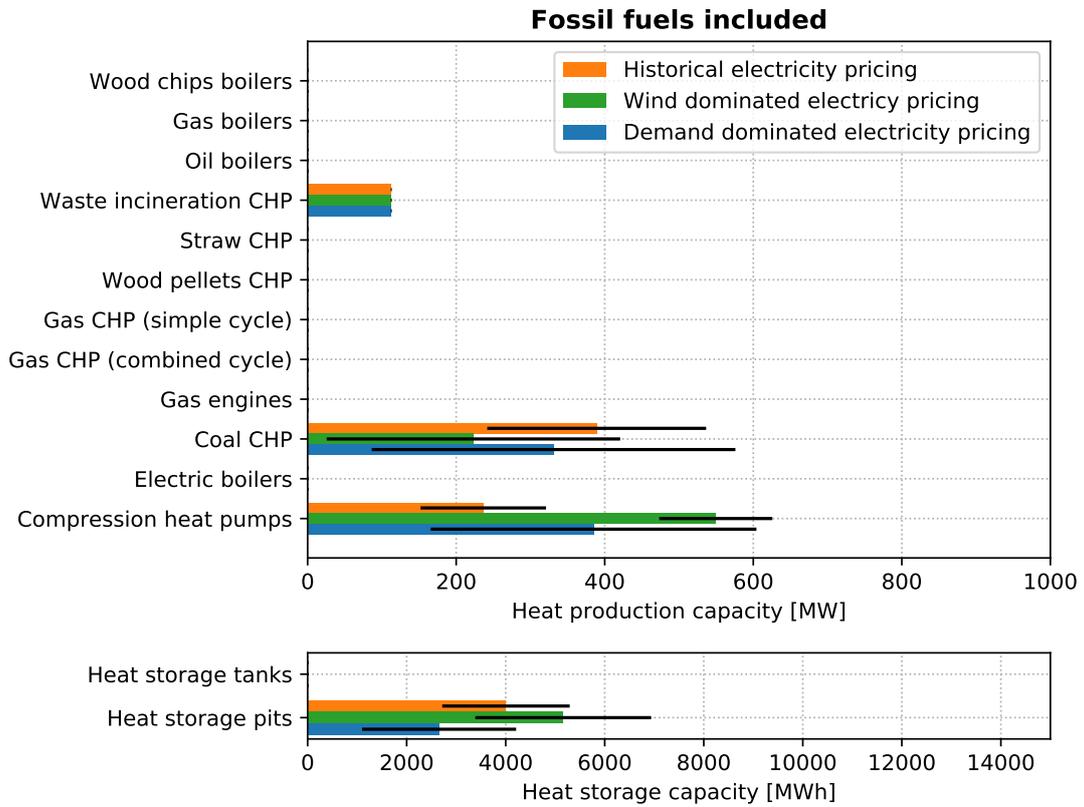
Table 3: Difference in the total annual Aarhus production system cost compared to the 2015 reference scenario. The cost differences are shown plus-minus an error of  $1\sigma$ . Negative values indicate lower than reference system cost.

Electricity pricing scheme	System cost difference [M€/yr]	
	Fossil fuels includes	Fossil free
Historical	$-8.4 \pm 3.8$	$-5.3 \pm 5.1$
Wind dominated	$-11.1 \pm 3.1$	$-9.7 \pm 4.6$
Demand dominated	$-6.7 \pm 3.9$	$-5.0 \pm 5.0$

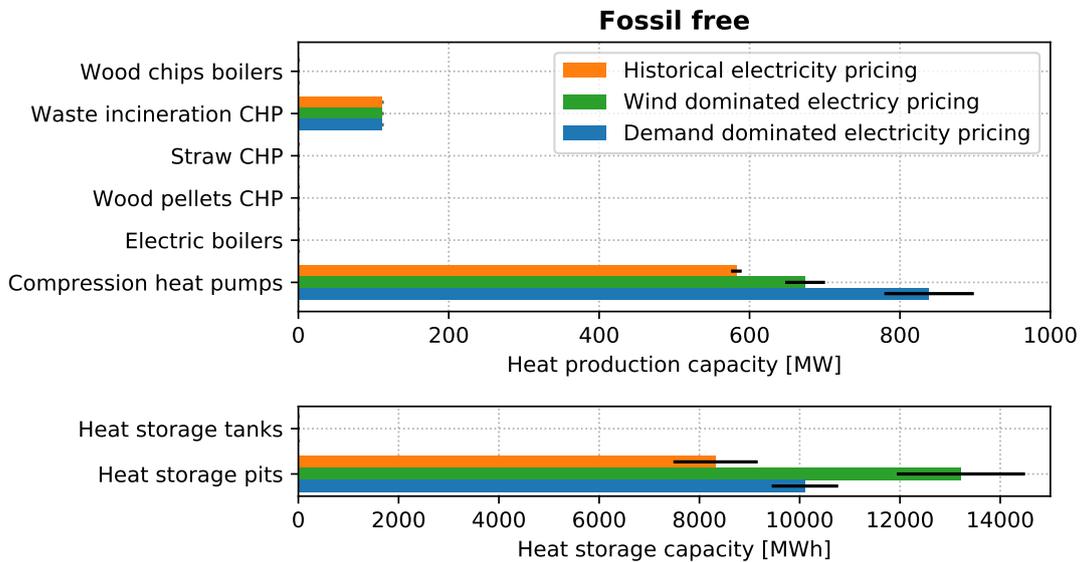
#### 3.1. The effect of the electricity pricing

As the wind power generation capacity in Northern Europe is expanding, it is likely that electricity prices in the future will become more strongly anticorrelated with the wind power production. In this study, we analyze how the heat production system is affected by a wind dominated electricity pricing scheme. An alternative scenario is also explored in which the electricity prices are dominated by the electricity demand instead of the wind power generation. The optimal heat production and storage capacities for these two electricity pricing schemes are shown as green and blue bars in Figure 3.

Wind power generation in Northern Europe has a positive seasonal correlation with the heat demand in district heating systems, because average winds tend to be higher in winter when it is also cold. This effect shifts the optimal heat production capacities toward larger shares of power-to-heat technologies and lower shares of CHP. Electricity becoming cheaper in the winter when the heat demand is high negatively impacts the economy of CHP units, while it benefits power-to-heat technologies. In the fossil fuel scenario, the optimal coal CHP capacity is almost halved



(a) Fossil fuels scenario.



(b) Fossil free scenario.

Figure 3: Optimal heat production capacity under three different electricity pricing schemes: Historical, wind dominated and demand dominated. In the top figure, all technologies are allowed in the optimization. In the bottom figure, only fossil free technologies are included. Error bars based on the sensitivity analysis are shown in black. The waste incineration capacity was fixed in the optimization.

while the heat pump capacity is more than doubled. In the fossil free scenario, the optimal heat pump capacity is also increased, but not as dramatically. The optimal heat storage capacity is increased regardless of whether or not fossil fuels are included.

The effect of implementing a demand dominated electricity pricing scheme depends on whether or not fossil fuels are allowed. If fossil fuel technologies are included, the demand dominated electricity price reduces the optimal coal CHP capacity and increases the heat pump capacity. The effect is similar to the effect of a wind dominated pricing scheme, although not as strong. The main difference is that in the demand dominated pricing scheme the storage need falls significantly instead of rising. In the fossil free scenario, the picture is different. The need for heat pumps increases significantly and so does the need for heat storage, although not as much as in the wind dominated scheme.

All in all, a future in which electricity prices are dominated by wind power production or by electricity demand is likely to increase value of heat pumps in the energy system at the expense of CHP units. Wind dominated electricity prices are also likely to increase the benefit of heat storages.

### 3.2. Cost sensitivity analysis

In order to assess the robustness of the optimal capacity configurations and system cost, a thorough sensitivity analysis has been performed. Using Latin hypercube sampling, we have run the system optimization in 200 perturbations of the initial values of fuel cost, investment cost and electricity price. The optimal capacities resulting from these cost-perturbed scenarios, can be seen on the radar charts in Figure 4. All CHP and boiler technologies have been aggregated to *Fuel based production capacity* and shown on the top axis. Each cost-perturbed scenario is shown as a triangle, plotted with an alpha transparency value, so 10 lines on top of each other appear as the full color. The initial unperturbed scenario is plotted as a red triangle corresponding to the orange bars in Figure 3. Notice that the scale is logarithmic and that production capacity is shown in units of MW, whereas storage capacity is shown in units of MWh.

The spread of the 200 triangles is an indication of the cost-based uncertainty in the optimal production and storage capacities. The wider the spread is, the larger the uncertainty is. Conversely, smaller spread indicates that the optimal configuration is robust to changes in investment, fuel and electricity costs. It is clear that the capacities in the fossil free scenario on the bottom are significantly more robust to changing costs than the fossil fuel scenario on the top.

It should be noted that 200 cost-perturbed scenarios have been solved for each of the other electricity pricing schemes, yielding very similar results. None of the cost-perturbations found it feasible to install capacities of dif-

ferent technologies than the ones that were assigned in Figure 3.

#### 3.2.1. Clustering in the optimal production system

Focusing on the top part of Figure 4, it appears that the capacities resulting from the different cost perturbations fall in different categories. Three clusters have been identified, when inspecting the data. Using the  $k$ -means clustering algorithm [35], we have assigned each resulting capacity configuration to one of the three clusters and colored them accordingly: green, blue and orange. An implementation of the algorithm from the Python framework scikit-learn (version 0.19.0) [36] was used.

Most of the perturbed cost-scenarios fall into the green cluster like the unperturbed scenario. The green main cluster consists of scenarios in which both heat pumps and coal CHP are installed in the production system in some mix. There is quite a bit of spread in this cluster and the heat pumps are installed with between 91 MW and 400 MW capacity, whereas the coal CHP is installed with between 345 MW and 1,080 MW. The anticorrelation between the power-to-heat capacity and the fuel based capacity can be observed from the crossover of the lines between the two vertices. The error bars assigned to the capacities in Figure 3a represent  $\pm 1\sigma$  where  $\sigma$  is the standard deviation of the capacity within the green main cluster.

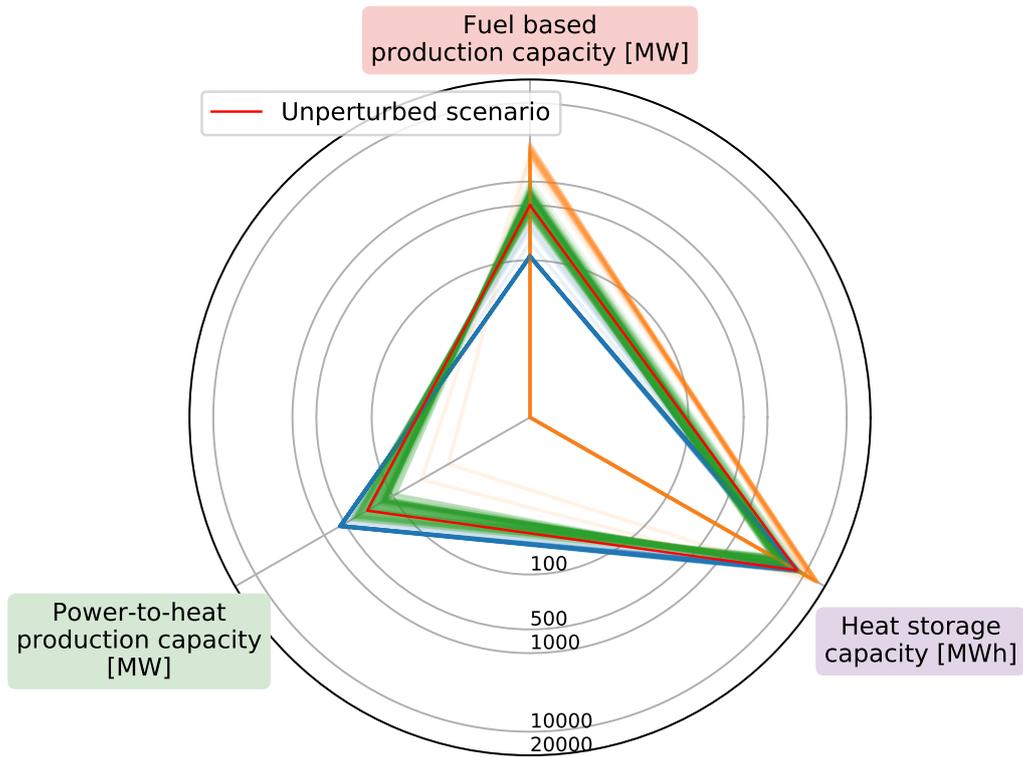
The blue and the orange clusters in the top of Figure 4 represent two opposite outcomes. The orange cluster contains all the scenarios in which the entire heat supply is covered by fuel based production: coal CHP and a little waste incineration. This group of scenarios is characterized by higher storage requirements and is mostly a result of the cost perturbations with significantly rising electricity prices.

The blue cluster is the opposite situation. In this group of cost scenarios, the city's heat demand is fully covered by power-to-heat technologies supplemented by a small base load of waste incineration. The storage needs in this cluster correspond to the high end of the storage capacity in the main green cluster.

Moving to the fossil free scenario on the bottom of Figure 4, all the cost perturbations fall into the same cluster: the blue cluster where power-to-heat technologies dominate the picture. The spread of the capacities in this cluster is quite narrow, as is also reflected by the error bars in Figure 3b.

Table 3 shows the reduction in system cost compared to the 2015 reference system. The cost are shown with an uncertainty of  $\pm 1\sigma$ , estimated within the main cluster. It is clear that going fossil free, the cost reductions are generally smaller and slightly more uncertain. The magnitude of the uncertainty makes it possible, that there may not be a cost reduction compared to the reference, especially in the demand dominated and in the historical electricity pricing scheme. The largest and most certain cost reduction would appear in the wind dominated electricity pricing scenario with fossil fuels allowed. The wider

## Fossil fuels included



## Fossil free

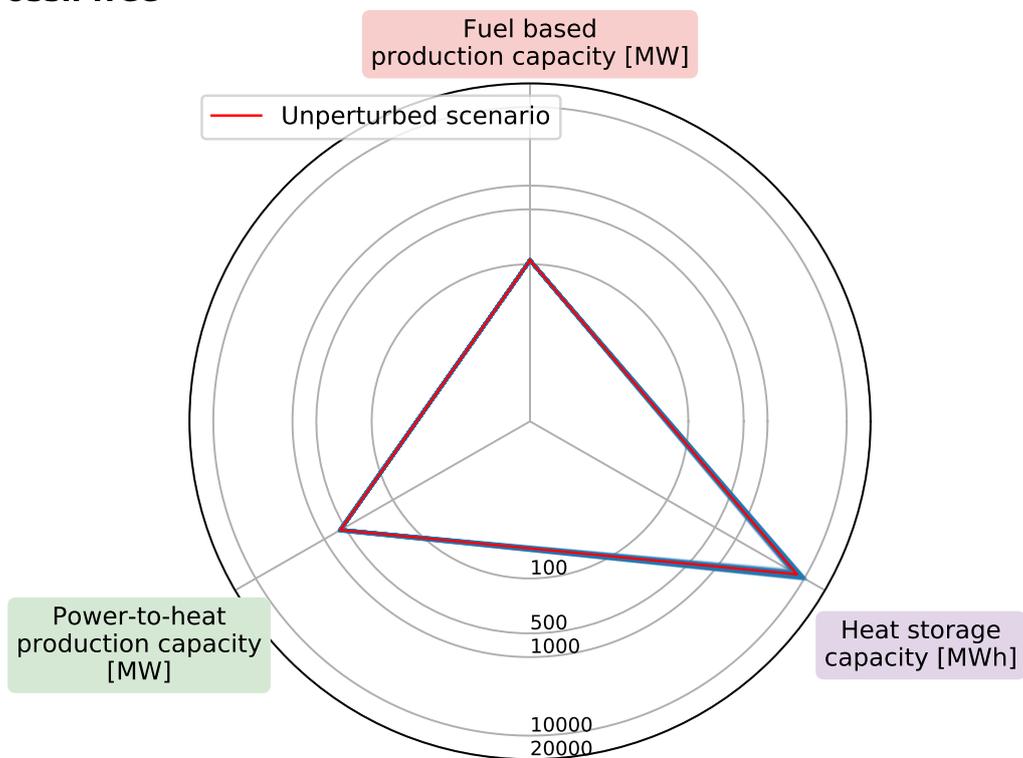


Figure 4: Radar charts of the optimal heat production capacities in 200 cost-perturbed sensitivity scenarios using the historical electricity pricing scheme. Fuel based capacity includes all boiler and CHP technologies, also waste incineration, which is not subject to the capacity optimization. In the top figure, all technologies are allowed; in the bottom figure, fossil fuel technologies, i.e. coal, oil and gas, are excluded. The unperturbed scenarios from Figure 3 (orange bars) are shown in red. The green, blue and orange triangles represent different clusters of the sensitivity scenarios. The scale is logarithmic, except the center, which represents 0. Production capacities are shown in units of MW and storage capacities are shown in units of MWh.

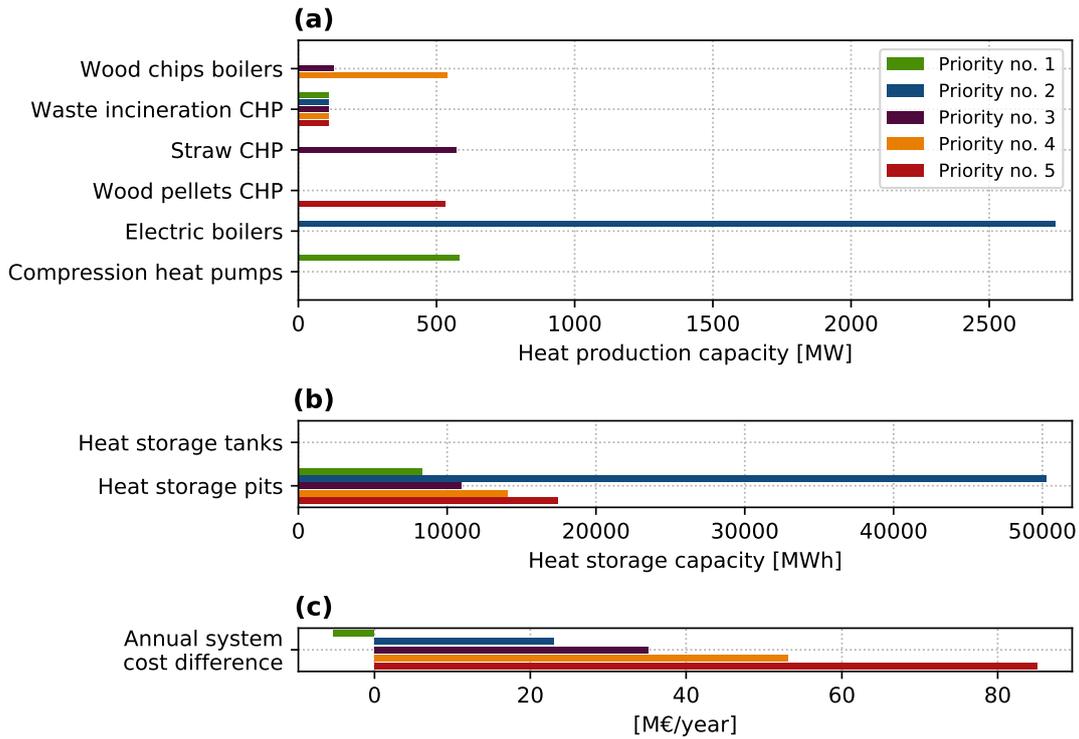


Figure 5: Ranking of fossil free production system alternatives, in case the first priority becomes infeasible. Each consecutive priority scenario is constructed by excluding the preferred production technology from the previous priority scenario. (a) shows the optimal heat production capacities and (b) shows the heat storage capacities. In (c) we display the additional annual system cost compared to the reference production system of 2015.

spread in the system cost in the fossil free scenarios, may be due to these scenarios being more vulnerable to rising electricity prices.

Summing up, when moving from away from fossil fuels, three different possible capacity configurations reduce to one and the spread of the optimal capacities narrows significantly. Heat pumps combined with storage and waste incineration remain as a highly robust choice for the future heat supply of the city.

### 3.3. Alternative paths to fossil free production

Since only very few technologies were installed in the optimal heat production systems, even under significant perturbations to the cost assumptions, we were interested in delving into different alternatives, in case the first choice of compression heat pumps became infeasible. There may be political roadblocks such as taxation of heat pumps or technical obstacles such as limited heat sources or weak electric grids, that make it infeasible to supply a whole city with heat from heat pumps. In order to find the second priority, in case heat pumps become infeasible we have optimized the production capacities and operation of a fossil free system excluding the heat pumps. Figure 5 shows the result of this analysis. The green bars show priority no. 1 and correspond to the orange bars in Figure 3b. This is the preferred fossil free scenarios, dominated by

compression heat pumps and with 8,322 MWh storage capacity. Excluding heat pumps from this scenario, brings us to priority no. 2.

In the second priority fossil free scenario, the heat pumps are replaced by a very large amount of electric boilers, and a very large amount of heat storage. The production system cost jumps significantly to be more than 20 million € more expensive per year compared to the reference system cost. This power-to-heat scenario, may be infeasible, due to the large areas needed for heat storage and the extreme number of electric boilers, putting a very high load on the local electric grid.

Excluding all power-to-heat technologies from the production capacity optimization brings us to priority no. 3. This is the first time we see biomass technologies enter the picture. In the third priority scenario, the heat for the city is provided primarily from straw CHP plants supplemented by a small amount of wood chips boilers and the waste incineration base load. This is a fossil free scenario that does not put a large strain on the electric grid and includes some dispatchable electricity generation. The necessary heat storage in this scenario drops to a more reasonable level, although still higher than in the first priority scenario. However, the cost of this system is even higher and about 150% of the cost of the reference system.

If straw CHP is not feasible on the scale needed for

priority no. 3, wood chips boilers completely take over the heat supply. In the fourth priority scenario, the heat is generated mostly from heat only technologies and the heat storage needs are higher than in the third priority. The total system cost in this case is more than 50 million € higher each year compared to the reference system.

Finally, if wood chips boilers are excluded from the optimization, the entire heat demand can be covered by wood pellets CHP plants in priority no. 5. While being able to serve as dispatchable backup for the electricity system, this would be an extremely expensive district heating production system. The system cost in this final scenario is now more than double the cost of the reference system. Note again that the system cost discussed here is not directly comparable to the cost of the present day Aarhus district heating system; rather, it is the cost of building a new system from scratch and operating it. The current system was built over many years, and transforming existing coal CHP plants into biomass CHP plants is cheaper than building entirely new plants.

It is clear that there are other ways to construct a fossil free district heating production system than by installing very large heat pumps and energy storage. However, all of these pathways require larger amounts of heat storage and increase the production system cost significantly.

#### 4. Conclusion and outlook

In this work, we have studied the cost-optimal production capacities in a city wide district heating system coupled to a larger electricity system. Using well-established technologies, i.e. heat only boilers, CHP units, power-to-heat technologies and heat storages, the optimal heat production has been characterized in a transition away from fossil fuels. The effects of electricity prices dominated by wind power production or by electricity demand have been investigated, and the uncertainty of the results have been estimated through extensive sensitivity analyses.

If we allow fossil fuels, the cost-optimal system will consist of a combination of coal CHP, heat pumps and heat storages. Going fossil free, heat pumps take over the heat supply, and the necessary storage capacity more than doubles, while the total system cost only increases slightly.

The optimal *choice* of technologies is highly stable under changing cost assumptions. But if fossil fuels are allowed, the optimal *capacities* of coal CHP and heat pumps are very uncertain. The need for heat pumps becomes significantly more certain if fossil fuels are banned. The total system cost, however, becomes more uncertain in the fossil free scenario, as it is more sensitive to changing electricity prices. A cost-optimal fossil free district heating system is thus more robust in its capacity allocation, but less robust in its cost.

There are other paths to fossil free district heating production, i.e. electric boilers or biomass, but these solutions all require larger heat storages and are significantly more costly.

Our study case, the Aarhus district heating system is going to change over the next 15 years, because key plants in the production system are at the end of their lifetime. This is an opportunity to rethink the production system, and our analysis indicates that regardless of a ban on fossil fuels, investing in large-scale heat pumps and heat storages is desirable, if taxes and regulations allow it.

Finally, the choice of technologies in this study was somewhat conservative, and only well-established dispatchable technologies were included. Future studies should include solar heating technologies, which may alter the system dynamics due to the seasonal and weather-dependent production patterns. New types of combined heat and electricity storages, e.g. the rock cavern storage described in [37], are emerging. Including combined heat and electricity storage technologies in the future may enhance the synergies between the electricity and heating sector in the transition away from fossil fuels.

#### Acknowledgements

This study is part of the READY project (Resource Efficient cities implementing ADvanced smart CitY solutions) which is partly financed by the EU's Research and Innovation funding program FP7 ([https://ec.europa.eu/research/fp7/index\\_en.cfm](https://ec.europa.eu/research/fp7/index_en.cfm)). We would like to thank AffaldVarme Aarhus for providing data about the heat load and production system in Aarhus. We also thank Energinet for providing data about electricity prices, consumption and power generation.

#### Appendix

The optimal operation and production capacities are found by solving a joint optimization problem. We pose the problem as a linear programming problem (LP) and minimize the total annual investment and operational cost. The objective function of the optimization problem reads:

$$\min_{\substack{\bar{P}_u^{\text{el}}, \bar{P}_u^{\text{heat}}, \bar{H}_s, \\ P_{u,t}^{\text{el}}, P_{u,t}^{\text{heat}}, \\ h_{s,t}, f_{s,t}}} \left( \sum_{u \in \text{prod. units}} c_u^{\text{el}} \bar{P}_u^{\text{el}} + c_u^{\text{heat}} \bar{P}_u^{\text{heat}} \right. \\ \left. + \sum_{s \in \text{storages}} c_s^{\text{stor}} \bar{H}_s \right. \\ \left. + \sum_{t=1}^N \Delta t \left[ \sum_{u \in \text{prod. units}} o_{u,t}^{\text{el}} P_{u,t}^{\text{el}} + o_{u,t}^{\text{heat}} P_{u,t}^{\text{heat}} \right. \right. \\ \left. \left. + \sum_{s \in \text{storages}} o_{s,t}^{\text{disp}} h_{s,t} + o_{s,t}^{\text{upt}} f_{s,t} \right] \right). \quad (1)$$

Here  $c$  denote annualized<sup>2</sup> capital cost per MW production capacity  $\bar{P}$  or per MWh heat storage capacity  $\bar{H}$ . The

<sup>2</sup>The capital cost was annualized using a discount rate of 4% and the lifetime listed in Table 2.

capital cost includes the nominal investment (CapEx) and fixed operation and maintenance cost (OpEx<sub>fixed</sub>).  $o_{u,t}^{\text{el}}$  and  $o_{u,t}^{\text{heat}}$  denote the marginal cost of electricity and heat production from unit  $u$  in hour  $t$ . The marginal cost includes fuel cost and variable operation and maintenance cost (OpEx<sub>variable</sub>). The optimization runs over an entire year, so the total number of time steps  $N$  is 8760 and the length of the time step  $\Delta t$  is 1 h. The rate of production of heat and electricity from unit  $u$  in hour  $t$  are denoted  $P_{u,t}^{\text{heat}}$  and  $P_{u,t}^{\text{el}}$ , and the rate of dispatch and uptake of heat from heat storage  $s$  we denote  $h_{s,t}$  and  $f_{s,t}$ , respectively. Finally,  $o_{s,t}^{\text{disp}}$  and  $o_{s,t}^{\text{upt}}$  are the marginal costs of dispatching from and storing heat in storage  $s$ .

### Constraints

The cost-optimization is imposed with a number of constraint, so the correct physical behavior of the system is captured.

*Energy balance.* The total electricity and heat load of the system must be met in all time steps  $t$ :

$$\sum_{u \in \text{prod. units}} P_{u,t}^{\text{el}} = P_{\text{tot},t}^{\text{el}}, \quad (2)$$

$$\sum_{u \in \text{prod. units}} P_{u,t}^{\text{heat}} = P_{\text{tot},t}^{\text{heat}}. \quad (3)$$

$P_{\text{tot},t}^{\text{el}}$  includes the total consumption in the local electricity market and any electricity consumed by power-to-heat technologies (see Figure 1).  $P_{\text{tot},t}^{\text{heat}}$  includes the total heat consumption in the city, as well as losses in heat storages and in the distribution system.

*Production capacity constraints.* Heat and electricity production are constrained by the production capacity for all units  $u$ :

$$0 \leq P_{u,t}^{\text{el}} \leq \bar{P}_u^{\text{el}}, \quad (4)$$

$$0 \leq P_{u,t}^{\text{heat}} \leq \bar{P}_u^{\text{heat}}. \quad (5)$$

*Storage constraints.* The energy content in the storage  $H_{s,t}$  is limited by the storage capacity:

$$0 \leq H_{s,t} \leq \bar{H}_s. \quad (6)$$

In any time step  $t$  the storage level  $H_{s,t}$  is governed by the dispatch and uptake of heat as well as the standing loss in the storage:

$$H_{s,t} = \eta_s^{\text{stand}} H_{s,t-1} + (f_{s,t} - h_{s,t}) \Delta t. \quad (7)$$

$\eta_s^{\text{stand}}$  is the standing heat loss factor. We also require cyclical storage operation, in order to avoid just depleting the storage in the end of the optimization period:

$$H_{s,t=1} = H_{s,t=N}. \quad (8)$$

We assume uptake and dispatch cost for the storages  $o_{s,t}^{\text{disp}}$  and  $o_{s,t}^{\text{upt}}$  to be 0.77 €/MWh, in order to counter excessive use of storages due to perfect foresight in the model. Except for the reproduction of the 2015 heat production, we have left the storage uptake and dispatch  $f_{s,t}$  and  $h_{s,t}$  unconstrained, as it depends on the installed pumping capacities. We have afterwards checked that the storage operation was sensible.

*Cogeneration constraints.* Our modeling includes two different types of CHP plants: An extraction-condensing plant (Type I) and a back-pressure plant with bypass (Type II). We adopt the notation from [29] and denote the power-to-heat ratio in back-pressure operation by  $\alpha_u$ . The specific electrical power loss, denoted by  $\zeta_u$ , is the extra heat that can be produced by reducing the electricity production by 1 unit while injecting the same amount of fuel [29].

*Type I.* For extraction-condensing plants, the electricity and heat production capacity are constrained by:

$$\bar{P}_u^{\text{heat}} = \frac{1}{\alpha_u + \zeta_u} \bar{P}_u^{\text{el}}. \quad (9)$$

Extraction-condensing plants are capable of running in condensing mode, where only electricity is produced. In a power versus heat diagram, the feasible operational area is below the top iso-fuel line

$$P_{u,t}^{\text{el}} \leq -\zeta_u P_{u,t}^{\text{heat}} + \bar{P}_u^{\text{el}}, \quad (10)$$

and above the back-pressure line

$$P_{u,t}^{\text{el}} \geq \alpha_u P_{u,t}^{\text{heat}}. \quad (11)$$

This is illustrated in Figure 6.

*Type II.* The other CHP type in the model is a back-pressure plant with bypass. This type of plant can bypass the steam turbine and boost the heat production by reducing the electricity production. It is assumed that 1 extra unit of heat can be produced for each unit of electricity not produced in bypass operation [27]. The total heat and electricity production capacities are thus constrained by:

$$\bar{P}_u^{\text{heat}} = \left(1 + \frac{1}{\alpha_u}\right) \bar{P}_u^{\text{el}}. \quad (12)$$

The feasible operational area for back-pressure plants with bypass in the power versus heat diagram is below the back-pressure line:

$$P_{u,t}^{\text{el}} \leq \alpha_u P_{u,t}^{\text{heat}}, \quad (13)$$

and below the bypass line

$$P_{u,t}^{\text{el}} \leq \bar{P}_u^{\text{heat}} - P_{u,t}^{\text{heat}}. \quad (14)$$

In this work we have modeled coal, wood pellets, gas engines and combined cycle gas CHPs as extraction-condensing Type I plants. Simple cycle gas, straw and waste incineration CHPs have been modeled as back-pressure Type II plants.

### Fuel consumption

The fuel consumption of boilers and CHP plants is governed by the efficiencies. For boilers, the fuel consumption for heat production is

$$P_{u,t}^{\text{fuel}} = \frac{1}{\eta_u^{\text{boiler}}} P_{u,t}^{\text{heat}}. \quad (15)$$

For CHP plants, the total fuel consumption for both heat and electricity consumption is

$$P_{u,t}^{\text{fuel}} = \frac{1}{\eta_u^{\text{el}}} (P_{u,t}^{\text{el}} + \zeta_u P_{u,t}^{\text{heat}}), \quad (16)$$

where  $\eta_u^{\text{el}}$  is the electrical efficiency of the plant.

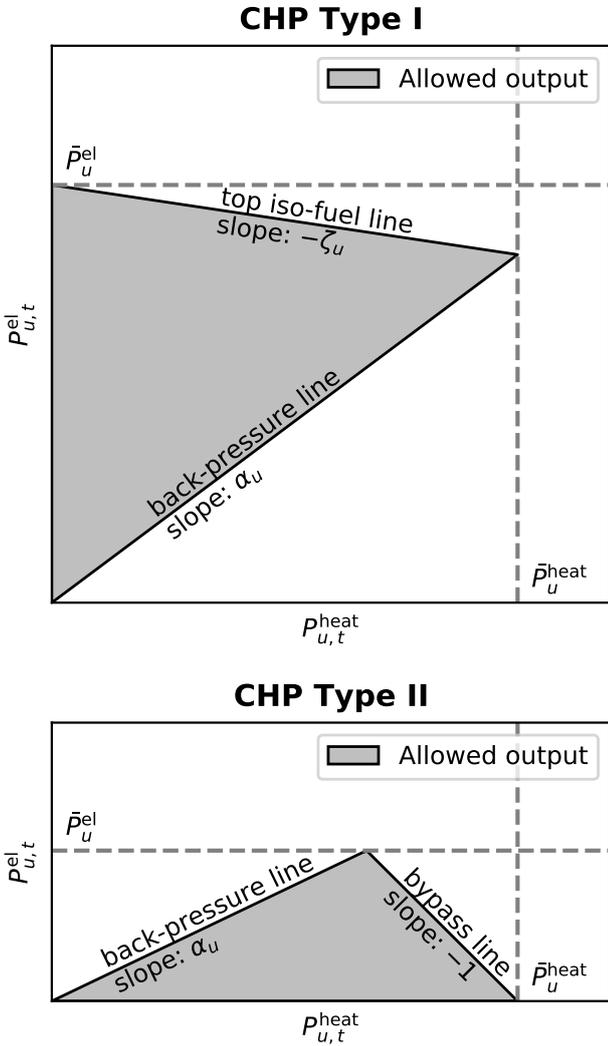


Figure 6: Feasible output of heat and electricity for the two CHP types included in the optimization. Type I is an extraction-condensing plant. Type II is a back-pressure plant with bypass.

### Nomenclature

$N$	number of time steps in the optimization
$\Delta t$	length of time steps in the optimization [h]
$\sigma$	standard deviation
$s$	index for storage units
$t$	index for hourly time steps
$u$	index for production units
$c_u^{\text{el}}$	annualized capital cost per MW electricity production capacity [€/MW]
$c_u^{\text{heat}}$	annualized capital cost per MW heat production capacity [€/MW]
$c_s^{\text{stor}}$	annualized capital cost per MWh heat storage capacity [€/MWh]
$o_{s,t}^{\text{disp}}$	marginal cost of dispatching heat from storage $s$ in hour $t$ [€/MWh]
$o_{u,t}^{\text{el}}$	marginal cost of electricity from unit $u$ in hour $t$ [€/MWh]
$o_{u,t}^{\text{heat}}$	marginal cost of heat from unit $u$ in hour $t$ [€/MWh]
$o_{s,t}^{\text{upt}}$	marginal cost of storing heat in storage $s$ in hour $t$ [€/MWh]
$P_{u,t}^{\text{el}}$	electricity production rate from unit $u$ in hour $t$ [MW]
$P_{u,t}^{\text{heat}}$	heat production rate from unit $u$ in hour $t$ [MW]
$h_{s,t}$	heat dispatch rate from storage $s$ in hour $t$ [MW]
$f_{s,t}$	heat uptake rate in storage $s$ in hour $t$ [MW]
$\bar{P}_u^{\text{el}}$	electricity production capacity of unit $u$ [MW]
$\bar{P}_u^{\text{heat}}$	heat production capacity of unit $u$ [MW]
$\bar{H}_s$	heat storage capacity of unit $s$ [MWh]
$P_{u,t}^{\text{fuel}}$	fuel consumption rate of unit $u$ in hour $t$ [MW]
$P_{\text{tot},t}^{\text{el}}$	total electricity load on the system in hour $t$ [MW]
$P_{\text{tot},t}^{\text{heat}}$	total heat load on the system in hour $t$ [MW]
$H_{s,t}$	heat content in storage $s$ in hour $t$ [MWh]
$\alpha_u$	power-to-heat ratio of CHP unit $u$ in back-pressure operation
$\zeta_u$	specific electrical power loss for CHP unit $u$
$\eta_u^{\text{boiler}}$	efficiency of boiler unit $u$
$\eta_u^{\text{el}}$	electrical efficiency of CHP unit $u$
$\eta_s^{\text{stand}}$	standing heat loss factor for storage $s$

### References

- [1] D. Jones, A. Sakhel, M. Buck, P. Graichen, The European Power Sector in 2017. State of Affairs and Review of Current Developments, Tech. rep., Agora Energiewende and Sandbag (2018).
- [2] C.-K. Woo, I. Horowitz, J. Moore, A. Pacheco, The impact of wind generation on the electricity spot-market price level and variance: The Texas experience, Energy Policy 39 (7) (2011) 3939–3944. doi:10.1016/j.enpol.2011.03.084.

- [3] H. Lund, F. Arler, P. A. Østergaard, F. Hvelplund, D. Connolly, B. V. Mathiesen, P. Karnøe, Simulation versus optimisation: Theoretical positions in energy system modelling, *Energies* 10 (7) (2017) 840. doi:10.3390/en10070840.
- [4] Z. Beihong, L. Weiding, An optimal sizing method for cogeneration plants, *Energy and Buildings* 38 (3) (2006) 189–195. doi:10.1016/j.enbuild.2005.05.009.
- [5] S. Gamou, R. Yokoyama, K. Ito, Optimal unit sizing of cogeneration systems in consideration of uncertain energy demands as continuous random variables, *Energy Conversion and Management* 43 (9) (2002) 1349–1361.
- [6] K. Sartor, S. Quoilin, P. Dewallef, Simulation and optimization of a CHP biomass plant and district heating network, *Applied Energy* 130 (2014) 474–483. doi:10.1016/j.apenergy.2014.01.097.
- [7] A. Rentizelas, I. Tatsiopoulos, A. Tolis, An optimization model for multi-biomass tri-generation energy supply, *Biomass and bioenergy* 33 (2) (2009) 223–233. doi:10.1016/j.biombioe.2008.05.008.
- [8] M. G. Nielsen, J. M. Morales, M. Zugno, T. E. Pedersen, H. Madsen, Economic valuation of heat pumps and electric boilers in the Danish energy system, *Applied Energy* 167 (2016) 189–200. doi:10.1016/j.apenergy.2015.08.115.
- [9] T.-M. Tveit, T. Savola, A. Gebremedhin, C.-J. Fogelholm, Multi-period MINLP model for optimising operation and structural changes to CHP plants in district heating networks with long-term thermal storage, *Energy Conversion and Management* 50 (3) (2009) 639–647. doi:10.1016/j.enconman.2008.10.010.
- [10] D. Romanchenko, J. Kensby, M. Odenberger, F. Johnson, Thermal energy storage in district heating: Centralised storage vs. storage in thermal inertia of buildings, *Energy Conversion and Management* 162 (2018) 26–38. doi:10.1016/j.enconman.2018.01.068.
- [11] J.-J. Wang, Y.-Y. Jing, C.-F. Zhang, Optimization of capacity and operation for CCHP system by genetic algorithm, *Applied Energy* 87 (4) (2010) 1325–1335.
- [12] M. Burer, K. Tanaka, D. Favrat, K. Yamada, Multi-criteria optimization of a district cogeneration plant integrating a solid oxide fuel cell–gas turbine combined cycle, heat pumps and chillers, *Energy* 28 (6) (2003) 497–518. doi:10.1016/s0360-5442(02)00161-5.
- [13] D. Buoro, P. Pinamonti, M. Reini, Optimization of a distributed cogeneration system with solar district heating, *Applied Energy* 124 (2014) 298–308. doi:10.1016/j.apenergy.2014.02.062.
- [14] D. Steen, M. Stadler, G. Cardoso, M. Groissböck, N. DeForest, C. Marnay, Modeling of thermal storage systems in MILP distributed energy resource models, *Applied Energy* 137 (2015) 782–792. doi:10.1016/j.apenergy.2014.07.036.
- [15] M. Åberg, D. Henning, Optimisation of a Swedish district heating system with reduced heat demand due to energy efficiency measures in residential buildings, *Energy Policy* 39 (12) (2011) 7839–7852. doi:10.1016/j.enpol.2011.09.031.
- [16] M. Münster, P. E. Morthorst, H. V. Larsen, L. Bregnbæk, J. Werling, H. H. Lindboe, H. Ravn, The role of district heating in the future Danish energy system, *Energy* 48 (1) (2012) 47–55. doi:10.1016/j.energy.2012.06.011.
- [17] A. N. Ünal, S. Ercan, G. Kayakutlu, Optimisation studies on tri-generation: a review, *International Journal of Energy Research* 39 (10) (2015) 1311–1334. doi:10.1002/er.3342.
- [18] M. Åberg, J. Widén, D. Henning, Sensitivity of district heating system operation to heat demand reductions and electricity price variations: A Swedish example, *Energy* 41 (1) (2012) 525–540. doi:10.1016/j.energy.2012.02.034.
- [19] V. Czitrom, One-factor-at-a-time versus designed experiments, *The American Statistician* 53 (2) (1999) 126–131. doi:10.2307/2685731.
- [20] M. D. McKay, R. J. Beckman, W. J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (2) (1979) 239–245. doi:10.2307/1271432.
- [21] M. A. Lozano, J. C. Ramos, L. M. Serra, Cost optimization of the design of CHCP (combined heat, cooling and power) systems under legal constraints, *Energy* 35 (2) (2010) 794–805. doi:10.1016/j.energy.2009.08.022.
- [22] T. Brown, J. Hörsch, D. Schlachtberger, PyPSA: Python for power system analysis, *Journal of Open Research Software* 6 (1). doi:10.5334/jors.188.
- [23] I. Gurobi Optimization, Gurobi optimizer reference manual (2016). URL <http://www.gurobi.com>
- [24] M. Dahl, A. Brun, G. B. Andresen, Decision rules for economic summer-shutdown of production units in large district heating systems, *Applied Energy* 208C (2017) 1128–1138. doi:10.1016/j.apenergy.2017.09.040.
- [25] Energiproducenttællingen for 2014–2016, Tech. rep., Danish Energy Agency (2017).
- [26] Technology data for energy plants: Generation of electricity and district heating, energy storage and energy carrier generation and conversion (May 2012), Tech. rep., Danish Energy Agency (Updated 2017). URL [https://ens.dk/sites/ens.dk/files/Analyser/technologydata\\_for\\_energy\\_plants\\_-\\_may\\_2012.\\_updated\\_2016\\_ver\\_nov\\_2017.pdf](https://ens.dk/sites/ens.dk/files/Analyser/technologydata_for_energy_plants_-_may_2012._updated_2016_ver_nov_2017.pdf)
- [27] Technology data for energy plants (August 2016), Tech. rep., Danish Energy Agency (Updated 2017). URL [https://ens.dk/sites/ens.dk/files/Analyser/technology\\_data\\_catalogue\\_for\\_energy\\_plants\\_-\\_aug\\_2016\\_update\\_oct\\_nov\\_2017.pdf](https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_energy_plants_-_aug_2016_update_oct_nov_2017.pdf)
- [28] Samfundsøkonomiske beregningsforudsætninger for energipriser og emissioner, Tech. rep., Danish Energy Agency (2017). URL [https://ens.dk/sites/ens.dk/files/Analyser/samfundsoekonomiske\\_beregningsforudsætninger\\_2017.pdf](https://ens.dk/sites/ens.dk/files/Analyser/samfundsoekonomiske_beregningsforudsætninger_2017.pdf)
- [29] S. Frederiksen, S. Werner, District heating and cooling, Studentlitteratur, 2013.
- [30] M. G. Rasmussen, G. B. Andresen, M. Greiner, Storage and balancing synergies in a fully or highly renewable pan-European power system, *Energy Policy* 51 (2012) 642–651. doi:10.1016/j.enpol.2012.09.009. URL <http://dx.doi.org/10.1016/j.enpol.2012.09.009>
- [31] H. Averfalk, P. Ingvarsson, U. Persson, M. Gong, S. Werner, Large heat pumps in Swedish district heating systems, *Renewable and Sustainable Energy Reviews* 79 (2017) 1275–1284. doi:10.1016/j.rser.2017.05.135.
- [32] Long term storage and solar district heating, Tech. rep., PlanEnergi (2016). URL [https://ens.dk/sites/ens.dk/files/Forskning\\_og\\_udvikling/sol\\_til\\_fjernvarme\\_brochure\\_endelig.pdf](https://ens.dk/sites/ens.dk/files/Forskning_og_udvikling/sol_til_fjernvarme_brochure_endelig.pdf)
- [33] S. Becker, R. A. Rodríguez, G. B. Andresen, S. Schramm, M. Greiner, Transmission grid extensions during the build-up of a fully renewable pan-European electricity supply, *Energy* 64 (2014) 404–418. doi:10.1016/j.energy.2013.10.010.
- [34] Energinet, Energi data service, Online. Accessed 2018.01.10. (Jan. 2018). URL <https://www.energidataservice.dk/>
- [35] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, Oakland, CA, USA, 1967, pp. 281–297.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [37] A. Arabkoohsar, G. B. Andresen, Design and analysis of the novel concept of high temperature heat and power storage, *Energy* 126 (2017) 21–33. doi:10.1016/j.energy.2017.03.001.



## Chapter 7

# Conclusion

I set out to explore economic and operational risk for planning and operation of district heating systems. Throughout the project, the focus has moved from short-term, over medium-term to long-term production planning. The gained insight has been used to provide tools to help production planners manage risk and make the best decisions in the face of uncertainty.

The weather is one of the most important sources of uncertainty affecting district heating production planning. Short-term production planning and system operation relies on heat load forecasts that use weather forecasts as input. I demonstrated how weather-based forecast uncertainty could be estimated, by introducing ensemble weather predictions into the field of heat load forecasting. The result was a probabilistic load forecast, allowing decision makers to play it safe when the forecast was less certain, and to push the system when the forecast uncertainty was small. As a first application case, I used the information about the time-dependent forecast uncertainties to reduce supply temperatures in the grid slightly, while maintaining security of supply.

Consumer behavior is another source of uncertainty affecting short-term heat load forecasts. The heat load on special occasions such as New Year's Eve is notoriously difficult to forecast accurately, because of the human component. I have augmented state-of-the-art heat load forecasts with data about local holidays, to better capture the effect of changing consumer behavior and improve forecast accuracy. The effects were small overall, but forecast performance improved somewhat on holidays and special occasions.

Moving on to medium-term or seasonal production planning, I quantified the economic risk related to the decision of shutting down production units over the summer. In a large district heating system, performing summer shutdown can have significant economic potential, but it is critical that the shutdown is timed accurately. Using an extensive weather dataset, I characterized optimal timing of these decisions with respect to seasonal weather patterns that vary greatly from year to year. I developed a set of decision rules for production planners and showed that they were capable of realizing most of the economic benefit from summer shutdown.

Finally, I turned to long-term production planning and investment decisions regarding future district heating systems. I assessed how uncertainties from variations in electricity prices, fuel prices and investment costs affected a cost-

optimal district heating production system. Installing only well-established production and storage technologies, large heat pumps and heat storages came to dominate the optimal system. If fossil fuels were allowed, coal CHP also had a role to play, but the ratio of coal CHP capacity to heat pump capacity was highly uncertain. It was found that banning heat production from fossil fuels would increase the system cost slightly, but it would drastically reduce the uncertainty on the optimal technology choice and capacity allocation.

The aim of this project was to provide insights about risk assessment in district heating that were scientifically sound, but with a short path to convert those insights to operational risk management tools for production planners. Two successful examples can be highlighted. The decision rules for the summer shutdown of a large CHP unit are ready for use in the Aarhus district heating system. Furthermore, the rules are straight-forward to adapt for application in other district heat systems. Likewise, the models developed for heat load forecasting in Chapter 4, are the foundation of a live forecasting system that is now in operation in the Aarhus district heating system. Without much overhead, this forecasting system can also be deployed to other district heating systems around the world.

# Bibliography

- Aarhus Kommune (2018, April). Befolkning. Web: <http://www.aarhus.dk/da/aarhus/Aarhus-i-tal/Befolkning.aspx>. Accessed 10.04.2018.
- Alessandrini, S., L. Delle Monache, S. Sperati, and G. Cervone (2015). An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy* 157, 95–110.
- Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.
- Arvastson, L. (2001). *Stochastic Modeling and Operational Optimization in District Heating Systems*. Centre for Mathematical Sciences, Lund University.
- Averfalk, H., P. Ingvarsson, U. Persson, M. Gong, and S. Werner (2017). Large heat pumps in Swedish district heating systems. *Renewable and Sustainable Energy Reviews* 79, 1275–1284.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Buoro, D., P. Pinamonti, and M. Reini (2014). Optimization of a distributed cogeneration system with solar district heating. *Applied Energy* 124, 298–308.
- Carpaneto, E., G. Chicco, P. Mancarella, and A. Russo (2011a). Cogeneration planning under uncertainty: Part I: Multiple time frame approach. *Applied Energy* 88(4), 1059–1067.
- Carpaneto, E., G. Chicco, P. Mancarella, and A. Russo (2011b). Cogeneration planning under uncertainty. part II: Decision theory-based assessment of planning alternatives. *Applied Energy* 88(4), 1075–1083.
- Connolly, D., H. Lund, B. V. Mathiesen, S. Werner, B. Möller, U. Persson, T. Boermans, D. Trier, P. A. Østergaard, and S. Nielsen (2014). Heat roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system. *Energy Policy* 65, 475–489.
- Czitrom, V. (1999). One-factor-at-a-time versus designed experiments. *The American Statistician* 53(2), 126–131.
- Dahl, M., A. Brun, and G. B. Andresen (2017a). Decision rules for economic summer-shutdown of production units in large district heating systems. *Applied Energy* 208C, 1128–1138.

- Dahl, M., A. Brun, and G. B. Andresen (2017b). Using ensemble weather predictions in district heating operation and load forecasting. *Applied Energy* 193, 455 – 465.
- Dahl, M., A. Brun, O. S. Kirsebom, and G. B. Andresen (2018). Improving short-term heat load forecasts with calendar and holiday data. *Energies* 11(7), 1–16.
- Dahl, M., R. A. Rodriguez, A. A. Søndergaard, T. Zeyer, G. B. Andresen, and M. Greiner (2017). Infrastructure estimates for a highly renewable global electricity grid. In *New Horizons in Fundamental Physics*, pp. 333–356. Springer.
- DeCarolis, J. F. (2011). Using modeling to generate alternatives (MGA) to expand our thinking on energy futures. *Energy Economics* 33(2), 145–152.
- Dotzauer, E. (2002). Simple model for prediction of loads in district-heating systems. *Applied Energy* 73(3-4), 277–284.
- Dotzauer, E. (2003). Experiences in mid-term planning of district heating systems. *Energy* 28(15), 1545–1555.
- European Commission (2011). Energy roadmap 2050. Technical report, European Commission. Available from: <http://ec.europa.eu>.
- Fang, T. and R. Lahdelma (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied Energy* 179, 544–552.
- Frederiksen, S. and S. Werner (2013). *District heating and cooling*. Studentlitteratur.
- Gneiting, T. and A. E. Raftery (2005). Weather forecasting with ensemble methods. *Science* 310(5746), 248–249.
- Grosswindhager, S., A. Voigt, and M. Kozek (2011). Online short-term forecast of system heat load in district heating networks. In *Proceedings of the 31st International Symposium on forecasting, Prague*. International Institute of Forecasters.
- Hyndman, R. J. and G. Athanasopoulos (2014). *Forecasting: principles and practice*. OTexts.
- Idowu, S., S. Saguna, C. Åhlund, and O. Schelén (2014). Forecasting heat load for smart district heating systems: A machine learning approach. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*.
- Inness, P. M. and S. Dorling (2012). *Operational weather forecasting*. John Wiley & Sons.
- Intergovernmental Panel on Climate Change (2015). *Climate change 2014: mitigation of climate change*, Volume 3. Cambridge University Press.

- ISO (2009). ISO/guide 73 risk management — vocabulary. *International Organization for Standardization, Geneva, Switzerland*.
- Izadyar, N., H. Ghadamian, H. C. Ong, Z. Moghadam, C. W. Tong, and S. Shamshirband (2015). Appraisal of the support vector machine to forecast residential heating demand for the district heating system based on the monthly overall natural gas consumption. *Energy 93*, 1558–1567.
- Jones, D., A. Sakhel, M. Buck, and P. Graichen (2018). The European Power Sector in 2017. State of Affairs and Review of Current Developments. Technical report, Agora Energiewende and Sandbag.
- Kato, K., M. Sakawa, K. Ishimaru, S. Ushiro, and T. Shibano (2008). Heat load prediction through recurrent neural network in district heating and cooling systems. In *Systems, Man and Cybernetics, 2008 IEEE International Conference on*, pp. 1401–1406. IEEE.
- Kristensen, M. H., R. Choudhary, R. H. Pedersen, and S. Petersen (2017). Bayesian calibration of residential building clusters using a single geometric building representation. In *Building Simulation 2017*, pp. 1650–1659.
- Lindenberger, D., T. Bruckner, H.-M. Groscurth, and R. Kümmel (2000). Optimization of solar district heating systems: seasonal storage, heat pumps, and cogeneration. *Energy 25*(7), 591–608.
- Lund, H., S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, and B. V. Mathiesen (2014). 4th generation district heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. *Energy 68*, 1–11.
- Madsen, H., K. Sejling, H. T. Sogaard, and O. P. Palsson (1994). On flow and supply temperature control in district heating systems. *Heat Recovery Systems and CHP 14*(6), 613–620.
- Marguerite, C., G. Andresen, and M. Dahl (2018). Multi-criteria analysis of storages integration and operation solutions into the district heating network of Aarhus—a simulation case study. *Energy 158*, 81–88.
- Mathiesen, B. V., H. Lund, and D. Connolly (2012). Limiting biomass consumption for heating in 100% renewable energy systems. *Energy 48*(1), 160–168.
- McKay, M. D., R. J. Beckman, and W. J. Conover (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*(2), 239–245.
- Möhrlen, C. and J. U. Jørgensen (2006). Forecasting wind power in high wind penetration markets using multi-scheme ensemble prediction methods. *Proceedings of DEWEK 2006*.
- Münster, M., P. E. Morthorst, H. V. Larsen, L. Bregnbæk, J. Werling, H. H. Lindboe, and H. Ravn (2012). The role of district heating in the future Danish energy system. *Energy 48*(1), 47–55.

- Nielsen, H. A. and H. Madsen (2006). Modelling the heat consumption in district heating systems using a grey-box approach. *Energy and Buildings* 38(1), 63–71.
- Ortiga, J., J. C. Bruno, and A. Coronas (2013). Operational optimisation of a complex trigeneration system connected to a district heating and cooling network. *Applied Thermal Engineering* 50(2), 1536–1542.
- Saha, S., S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M. H. Juang, J. Sela, M. Iredell, R. Treadon, D. Kleist, P. V. Delst, D. Keyser, J. Derber, M. Ek, J. Meng, H. Wei, R. Yang, S. Lord, H. van den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin, P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. W. Reynolds, G. Rutledge, and M. Goldberg (2010). NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010.
- Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H. ya Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen, and E. Becker (2011). NCEP Climate Forecast System Version 2 (CFSv2) Selected Hourly Time-Series Products.
- Schmidt, O., A. Hawkes, A. Gambhir, and I. Staffell (2017). The future cost of electrical energy storage based on experience rates. *Nature Energy* 2(8), 17110.
- Sculley, D., G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems*, pp. 2503–2511.
- Taylor, J. W. and R. Buizza (2002). Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems* 17(3), 626–632.
- Taylor, J. W. and R. Buizza (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting* 19(1), 57–70.
- Taylor, J. W., P. E. McSharry, and R. Buizza (2009). Wind power density forecasting using ensemble predictions and time series models. *IEEE Transactions on Energy Conversion* 24(3), 775–782.
- Traber, T. and C. Kemfert (2011). Gone with the wind?—electricity market prices and incentives to invest in thermal power plants under increasing wind energy supply. *Energy Economics* 33(2), 249–256.

- Werner, S. (1984). *The heat load in district heating systems*. Ph. D. thesis, Chalmers tekniska högskola.
- Woo, C.-K., I. Horowitz, J. Moore, and A. Pacheco (2011). The impact of wind generation on the electricity spot-market price level and variance: The Texas experience. *Energy Policy* 39(7), 3939–3944.