Multi-modelling for the energy transition

Exploring coupling-based issues in multi-resolution energy multi-models

B. Boereboom

MSc thesis

Cover image by Andreas (2019)

TU Delft

Multi-modelling for the energy transition

Exploring coupling-based issues in multi-resolution energy multi-models

by

B. Boereboom

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday August 26, 2022 at 13:30.

Student number:	4556232
Project duration:	February, 2022 – July, 2022
Thesis committee:	
Chair: Dr. ir. I. (Igor) Nikolic,	TU Delft, systems engineering
First supervisor: Dr. Y. (Yilin) Huang,	TU Delft, systems engineering
Second supervisor: Dr. ir. E.J.L (Emile) Chappin,	TU Delft, energy and industry

TU Delft

Preface

This document contains my master thesis and serves as the final work in my 5-year journey at the faculty of Technology, Policy and Management at the TU Delft. I started this study program, BSc and MSc, to gain the skills and knowledge required to be able to do my part in realising the energy transition. During the years at TPM I regularly found myself fascinated and enthusiastic about the variety of subjects that I was gratefully able to follow, in combination with the friendly and open environment that the subjects were taught in. I would like to thank my teachers and supervisors Yilin Huang, Igor Nikolic and Emile Chappin for inspiring me during their lectures and for guiding me throughout the thesis process.

I would also like to thank my good friends Lieuwe Berdowski, Floris Boendermaker, Bruno Hermans and Wesley Nijmeijer, who took this TPM adventure with me. I met them during the first week of the bachelor due to random group assignment from a computer. We became the best of friends and it is thanks to their friendship and support that I was able to complete many projects, the BSc thesis and now this master thesis.

Bram Boereboom
Delft, July 2022

Summary

The Dutch government described its vision to achieve significant reductions in greenhouse emissions in the climate goals of 2030 and 2050. The energy infrastructure in the Netherlands will be a critical factor in achieving these goals. As an extension to the stated importance, understanding the current energy infrastructure is equally significant. Models often form the base of understanding such complex systems. Unfortunately, the modelling environment of the Dutch energy infrastructure is fragmented. There exists no one model that is able to give comprehensive oversight in order to provide policy makers with the information needed to facilitate key energy policy decisions. However, there is a variety of models present that each clarify their own piece of the puzzle. Therefore it is the ambition of the TU Delft and partners to create a multi-model infrastructure that is able to couple existing energy models in order to facilitate comprehensive energy policy creation. This thesis is part of the research needed in order to achieve this higher goal.

The research documented in this thesis is focused on the problem of detecting and alleviating coupling-based issues that emerge from coupling energy models operating on different resolutions, that weren't originally built to be coupled. Many models differ, slightly or strongly, on the level of detail on which they operate. Such a difference is called a resolution-based difference. Differences could be a difference in time detail such as the timestep size, in spatial detail or in the detail of which the behaviour of objects are modelled. In order to couple models with differing resolutions, the gap between the resolutions must be bridged. However, bridging the gap can bring along a variety of issues. Detecting which issues are present in a coupling effort and how to effectively alleviate them remains difficult. The research question of this thesis is therefore as follows:

How can issues that arise when coupling multiple energy models that have different resolutions be resolved effectively?

The main conclusion of the conducted research is that *the problem described in the main research question can be answered by using a coupling process based on audits, comprised of questions aimed at detecting issues and checking the effectivity of the means to solve the issues* (see figure 1). These audits have proven successful in completing two different multi-resolution multi-modelling coupling case studies. The process entails (for a two-model coupling) a separate model audit for each model, leading to a coupling audit using both models and finally to the realisation of the coupling itself. Adherence to the process described in figure 1 standardises the way couplings are created to a degree. Because of this, model audits done for one coupling could for example be re-used at a later date for another one. This provides value over old coupling methods, which were often done individually in an ad-hoc manner.

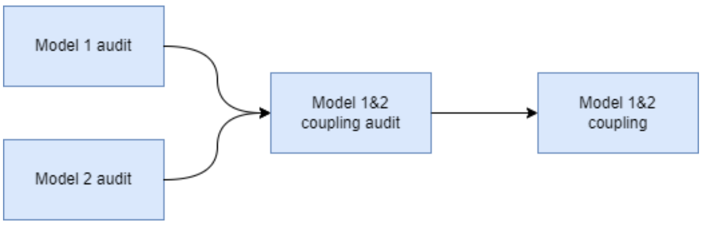

Figure 1: Graphic representation of audit method

The audits were found to facilitate the process of getting to know the models in the coupling, as well as identify the specific facets that should be taken into account when deciding on proper (dis)aggregation functions and consistency checks. In addition, the audits were expanded whenever issues in the coupling were located that the iteration of the audits at the time was unequipped to detect. Out of the two challenges that stem from the main research question, recognising issues first and creating methods to alleviate them second, the research concludes that *the first challenge proves to be*

the more laboursome of the two. Recognising which issues were present in the case studies required a deep dive into the models, guided by the model audit. The lessons learned there can be brought over to the coupling audit and the eventual coupling functions themselves. Once the audits were done, the actual creation of proper coupling functions was not a major obstacle.

Additionally, it was found that *the coupling of two models will likely require the construction of an A/D effort specifically tailored to the needs of the models in question.* Due to the nuances and subtle differences present in model couplings, it is unlikely that (dis)aggregation functions made for one coupling can be directly transferred to another one. The case studies showed that some parts of the time disaggregation functions of case study one could be re-used in a similar time disaggregation in case study two. However, differences in the specific time disaggregation and the consistency requirements necessary still requires substantial adaptation of the functions.

Furthermore, *system expertise is highly advantageous for identifying and understanding possible problems during the auditing and coupling process.* Especially the coupling audit benefits from system expertise, as it makes checking semantic overlap easier. Energy modelling was chosen as a subject for the coupling case studies, partially because energy systems are the field of study of the researcher. It is expected that this need for system expertise will also prove useful when coupling models in other fields such as finance or logistics. However, further research is necessary to confirm this.

Finally, the research concludes that *such couplings that require an A/D effort need a data model in the middle to translate (and possibly store) data to facilitate the coupling.* A modeller coupling two such models as in the case studies must be aware that inserting a data model introduces a third model into the mix. This additional (data)model can be prone to biases which might influence the results of the model coupling.

The proposed methods are not without their limits. A limiting factor of the audits is the absence of hard thresholds on what constitutes good consistency or sufficient semantic overlap. The audits contain questions about inquiring information on these topics, but provide no benchmark numbers to compare results to. Furthermore, the audits cannot guarantee that all typical issues present in a coupling effort will be detected. The presence of biases in the questions themselves, as well as the applicability of the auditing method on more dynamic or non energy-based models remains an unexplored subject. It is therefore recommended in chapter 7 that further research can focus on the mitigation of the limiting factors described, or on the exploration of the avenues of research currently left unknown.
The research done to arrive at the conclusions was organised into the following chapters:

- Chapter 1 further clarifies the motivation behind the research objective. The main research question is defined, along with several sub-questions in order to split the problem into more concrete steps.
- Chapter 2 contains an extensive literature review on all manner of relevant fields to the research questions. Literature on multi-modelling, multi-resolution modelling, aggregation and disaggregation methods and consistency maintenance is reviewed and documented.
- Chapter 3 uses these categories of possible issues to create two audits: a model audit and a coupling audit. Both audits are lists of questions that are aimed to extract specific information from the to be coupled models, in order to detect possible issues present. The two audits are meant to be used to facilitate the coupling process as described in figure 1. To test and improve the audits, two case studies containing hypothetical couplings are set up and described. The case studies both involve coupling two energy models that function on differing resolutions.
- Chapter 4 contains the model- and coupling auditing of the models present in both case studies.
- Chapter 5 contains the final coupling of both case studies. (Dis)aggregation functions are created using the insights gained from the audits in chapter 4. After the coupling has been completed, a validation is performed on the (dis)aggregation functions used.
- Chapter 6 contains reflection on the case studies themselves and a general discussion of the effects, limitations and unknown factors of the auditing method.
- Chapter 7 provides an oversight of the conclusions drawn from the research, as well as recommendations for future research as described above.

Contents

1 Introduction

1.1. Introduction

The current Dutch energy infrastructure is old, multifaceted and incredibly complex. Technical, economic, environmental, societal and political forces are all trying to influence this infrastructure to move in a certain direction. An important fact is that this current energy infrastructure will have to be used to attain our new climate goals (TenneT, 2021a). Therefore it is paramount to solidify our understanding of the Dutch energy infrastructure in order to more easily reach the 2030 & 2050 climate goals (EZK, 2019).

Every facet of the chain is governed by actors with their own models regarding how their segment functions. This includes models such as production, transportation, load balancing, distribution and market trading. Unfortunately, at this level of detail, there exists no single model that oversees an entire chain. Models used for energy policy purposes are either at a higher abstraction level or do not contain the entire system being studied. Separated modelling and piece-wise understanding of the system is not sufficient when studying complex socio-technical systems such as this, because the it is the behaviour of the whole system that ought to be studied. This could be more than the sum of the behaviour of its parts due to potential interaction effects (Vangheluwe et al., 2002).

Creating one all-encompassing model would be a possible solution. However, the sheer complexity and diversity of the systems makes this difficult. Some systems, like voltage across electricity grids, better fit continuous models, while industrial processes like coal plants might favour a discrete or process-based approach. When the broadness of a system is not able to be (efficiently) covered in one model (with one modelling formalism), one could use a multi-model instead as a better fit (Bankes, 2002). Multi-models are a set of models interacting together. This approach would be able to capture all specific facets of a chain in an efficient way, due to being able to use multiple formalisms simultaneously (Yilmaz et al., 2007). In addition to this, multi-modelling has the benefit of preserving previous investments done in the singular models (Brandmeyer & Karimi, 2000).

This thesis is part of a larger project that aims to create a multi-model infrastructure to couple existing models from chain partners into one platform, in order to attain better understanding and create more detailed policy (Multi-model.nl, 2021). The target is to couple between 5-18 models in this platform. One challenge that needs to be resolved before this infrastructure can function is how to deal with any issues related to the coupling of models created on different resolution levels (Nikolic et al., 2019). The scope of this thesis will therefore be exploring which issues stemming from coupling multiple-resolution models exist and how they can be alleviated. Briefly, model resolution concerns itself with the granularity and aggregation level of models. Higher models generally have the benefit of being able to capture more detail. A Lower model resolution also has its advantages, such as when models only need to provide a general overview of something, instead of a detailed picture. Lowering the model resolution enables a faster model that can still deliver the needed product.

1.2. Knowledge gap and research questions

This section covers the knowledge gaps that will be addressed by the thesis. In order to be concise, the origins of the presented gaps will be motivated briefly for each question. The literature review in chapter 2 will go into depth on where the gaps originate from. The methodology of how each research question will be answered is documented separately in section 1.3.

1.2.1. Main research question

Multi-modelling has previously not been used in the Dutch energy sector. It offers the ability to create an integral and all-encompassing framework of knowledge with which policy makers can make decisions and see the impacts of the whole system at once, not bit by bit. Using existing models provides a multi-model infrastructure with detailed models. These will be based on the stakeholders' own understanding of their own part of the system, instead of an outside modellers' view on it. However, combining and incorporating models into each other that were not built for this brings along several pitfalls, of which resolution-based issues will form the focus of this thesis. Therefore the main research question reads as follows:

How can issues that arise when coupling multiple energy models that have different resolutions be resolved effectively?

1.2.2. Sub-questions

To more gradually answer the main research question, the problem will be broken down and answered in pieces with several sub-questions (SQ's).

SQ1

A natural starting point for answering the main research question is to explore what issues can arise when coupling models with differing resolutions. Sub-question 1 is aimed at exactly that:

What are typical issues that arise when coupling multiple energy models that have different resolutions?

SQ2

After the typical issues stemming from literature have been identified, the question shifts to how well the findings in literature translate to the case of specifically issues when coupling energy models. A case study is required featuring two energy models, differing in resolution, that are suitable for analysis. A suitable case study will need to offer enough "struggle" in coupling, alongside a barrier in the form of differing resolutions that need to be bridged. In short, the selection of an appropriate case study is an objective of its own. The second research question is dedicated to this objective:

What is a suitable case study aimed to analyse these issues?

SQ3

Once the models for the case study have been selected, resolution-based coupling issues can be analysed with them. At the start of the case study it is unknown which subset of issues, from the relatively broad selection of possible issues documented in literature, are present in the case study. Furthermore, it is unknown how one can extract from a given model which possible issues there are. This knowledge gap will have to be addressed before any attempts are made at creating methods to alleviate the issues. Therefore, research question 3 is as follows:

How can the presence of typical issues be recognised?

SQ4

The answers from SQ3 provide a way to identify the issues present in the models in the case study. Once a list of issues has been created, the next step is to create methods to alleviate these issues. This step is also where the actual coupling of the two models will take place. The coupling has to be done using methods designed to alleviate the identified coupling issues. What these methods are and how to apply them to the case is the knowledge gap that is addressed by sub-question 4:

What methods can be used to alleviate these issues?

SQ5
In this step, the methods have been implemented and the models have been coupled. However, the effectiveness of the methods has yet to be analysed. This step is aimed at identifying what the effects and limits are of utilising these issue-alleviating methods is on the model coupling.

What are the effects and limitations of these methods?

1.3. Methodology

The main research question has been split into five sub-questions to lay out the process steps in which the main question will be answered. A methodology is presented on how each sub-question can be solved, alongside their desired outputs.

1.3.1. SQ1 methodology

A literature study can be used to answer this question. Search engines such as Scopus and Google scholar can be used for the creation of a literature list that ideally contains a diversified range of papers from different publishing years, countries and perspectives.

The desired output of asking sub-question 1 is to gain a thorough understanding of the literature surrounding the topic. Key terms like coupling, model resolution and consistency will be defined in this section of the thesis. Papers found during this phase will also prove useful in answering the other sub-questions

1.3.2. SQ2 methodology

Subquestion 1 exposes a general view of problems that could arise. With this knowledge in mind, a case study can be selected that is sufficiently "tough" to couple to properly analyse the typical issues. An ideal case study would be two models built in different manners for different purposes that need to be coupled.
For this thesis, the scope of models that can be selected for the case study will be limited to python models. This is due to python being the language that the researcher is most experienced in. Python will also be the language used to construct the appropriate mapping functions (methods) and consistency-checking functions to couple the models.
N.B. The aim is not to create a real-time coupling, as this brings with it a significantly more difficult coding effort. Real-time coupling is also not the focus of this thesis. The coupling method itself, and whether it is addresses the typical issues, is the focus.

The desired output of SQ2 is a case study suitable for further analysis and testing. The selection of the case study was specifically chosen to be at this point (SQ2) of the thesis, because a practical example can be used to test, explore and iteratively improve findings from SQ3-5.

1.3.3. SQ3 methodology

Before any methods for resolving coupling issues in the case study can be created, first the issues have to be recognized. The method used to recognize possible issues in the case models is to create a list of questions that will 'audit' the models. The list(s) will be created by first examining the literature in chapter 2 to see how previous researchers recognized possible issues. From that point, possible problem-causing areas in the models can be identified. These problem areas will then be included in the auditing list.

The case study can help in improving this auditing list through iteratively checking whether the list provides sufficient coverage to find the issues present in the case study. If during any of the following sub-questions it becomes apparent that parts of the auditing list are not present or redundant, it can again be iteratively improved.

The desired output of SQ3 is an auditing list of questions, used to recognize which typical issues are present in the models at hand. Due to the timescope of this thesis, it will be difficult to explore whether the auditing list is sufficient enough to use on recognizing issues in models other than the case study.

Whether and how the auditing list could be used to generally identify resolution-related coupling issues will be addressed in the discussion.

1.3.4. SQ4 methodology

The models in the case study paired with insights from literature will be used in order to create problem-alleviating methods. These methods will take the form of mapping functions and consistency checking functions in python.
In creating these methods, there are two main schools of thought:

1. Methods to alleviate resolution-based issues that originate from inside the individual models, meaning that the coupling issues arise from the characteristics of the model itself.
2. Methods to alleviate resolution-based issues that originate from the coupling effort between the models. This refers specifically to the way in- and outputs are being mapped that could cause issues.

By “coupling issues arising from the characteristics of the model itself” it’s referring to the specific way that the model is built and how that can influence how the model should be coupled. Examples of such characteristics include the examples given in the methodology of SQ3

It is important to note that, just like problem recognition, problem alleviation through the creation of methods is an iterative process that will require a lot of going back and forth between the methods, the case and the literature.

The desired outputs of SQ4 are:

- A python file of mapping functions and consistency-checking functions that came forth as an implementation of how the recognized typical issues in the models can be solved.
- A functional coupling between the two models in the case study that uses the aforementioned functions to communicate between the models.

1.3.5. SQ5 methodology

After the couplings in the case studies are completed, the results will be analysed. Especially the ability of the method to alleviate the typical issues that were recognised will be investigated.

The desired output of SQ5 is a clear overview on the effectivity and limits of the described methods to alleviate resolution-based coupling challenges. Knowledge regarding where the described methods were unable to alleviate the problems are just as insightful as knowledge of more successful methods. Knowing what problems have yet to be solved is a great indicator for where future research could be done.

1.4. Relevance to MSc program

The associated MSc for this thesis is Engineering and Policy Analysis (EPA) at the TU Delft, the Netherlands. The EPA MSc focusses on analysis complex socio-technical systems. It incorporates modelling, policy and data-analysis in its mission of solving grand challenges such as sustainability (TU Delft, 2022). This thesis covers the topics of simulation and multi-modelling in the energy sector, in order to contribute to a better policy advising tool for the energy transition. The subject matter, as well as the goal of the thesis, are well aligned with those of the EPA MSc. The thesis allows for the use of many techniques that are taught in EPA courses such as (data) modelling, reviewing of scientific literature and aiding in the effort of resolving a grand challenge (energy).

1.5. Thesis structure graphic

Figure 1.1 shows how the thesis will be structured, alongside the chapters that the various parts of the research will be documented in.

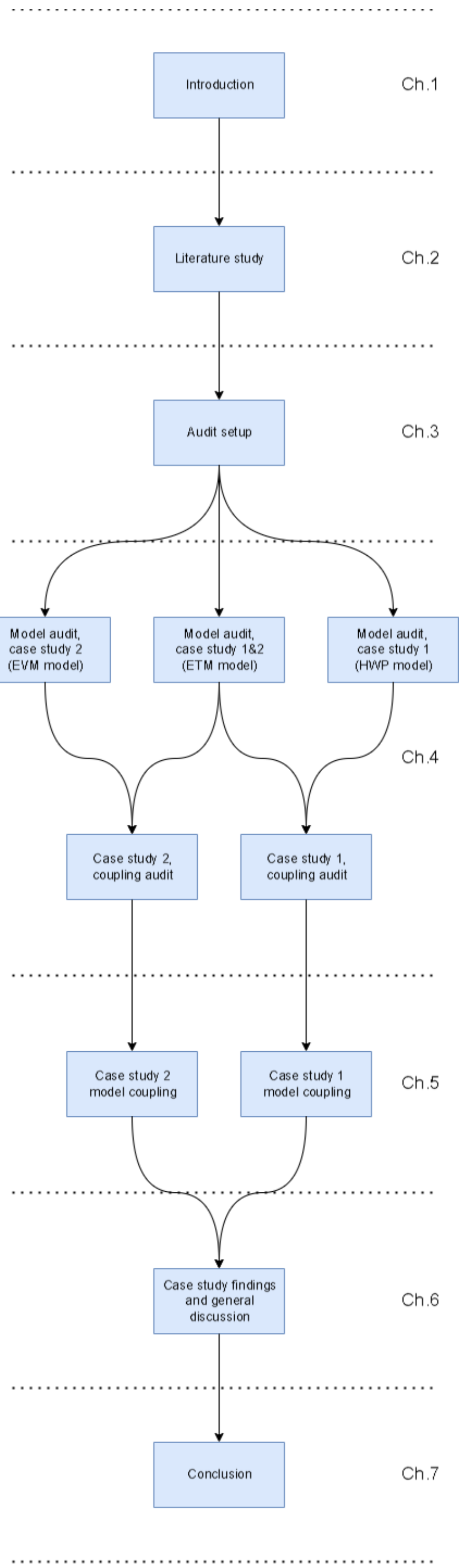

Figure 1.1: Graphic representation of the thesis structure

2
Literature review

In order to gain an in depth understanding of both the current knowledge regarding multi-resolution multi-modelling, as well as literature on issues relating to resolution difference in modelling, a literature review was conducted. Key terms such as multi-modelling, multi-resolution modelling and consistency are defined and discussed. While laying out the literature, special attention shall be given to any resolution-related coupling problems that previous researchers have encountered.

2.1. Literature selection method

Literature was sourced from online search engines such as Scopus and Google Scholar.

2.2. Multi-modelling

The variety between processes in the energy infrastructure shows in the way that these sectors are being modelled. Electricity is a continuous product, thus requiring continuous models. Industrial processes often contain mixes of discrete and continuous processes. For example, the discrete insertion of raw materials (coal) into a continuous reaction process (power generation). Because of this distinct variety in the functioning of energy infrastructure, the way that they can be modelled varies as well. To facilitate these systems in transitioning towards the climate goals, many visions have been created, such as the exploration of future energy infrastructure (TenneT, 2021b), the Dutch national program of regional transitions (RES, 2021), and the heat transition reports (RVO, 2021). These reports all work on improving their specific sections, utilizing their specific modelling methods. However, no integral and broad spanning tool of modelling the full impact of these changed has been implemented. In order to come to a more integral understanding of the energy infrastructure systems in the Netherlands, a multi-model infrastructure that incorporates models from all facets of energy could be developed (Multi-model.nl, 2021).

2.2.1. Multi-modelling use-case

Multi-modelling is a relatively novel technique that utilises multiple models, each with their own interpreter, on one network to run a model. In contrast to parallel simulation which runs multiple independent simulations on different processor cores, multi-modelling works on one simulation at once through multiple models influencing each other (Perumalla, 2006). Multi-modelling as a tool for the energy system was chosen for a few reasons:

- Modellers are able to use existing models. You don't have to reinvent the wheel when multi-modelling to the extent that one would while making a whole new model. Multi-modelling can save time and money due to not having to create new models. In addition, multi-modelling also makes sure that previous company assets (models) aren't discarded, thus retaining the investment (Brandmeyer & Karimi, 2000).
- The resources (models) used can be geographically distributed, which is a big advantage when dealing with spatially and digitally separated machines and models (Fujimoto, 2015).

- Using in multiple models built by a variation of people brings a variety in perspectives on the system, which can expand and improve the view on the studied system (Seck & Honig, 2012).
- Privacy issues between separate for-profit companies can restrict their willingness to give out their entire model. Multi-modelling enables actors to keep their models in-house while simultaneously enabling communication.

In multi-modelling, the models to be coupled might have differing resolutions. Before a coupling can be made, the differences in resolution must be addressed, either by the creation of intermediary conversion algorithms (Brandmeyer & Karimi, 2000; Tan et al., 2001) or via the conversion of the original models to run on the same resolution (Natrajan et al., 1997; Reynolds et al., 1997). Out of the two options, intermediary conversion algorithms serve the purposes of the multi-model infrastructure better. Conversion of existing models to a different resolution is a time-intensive process, although it can be done (Salome, 2021). However, this would forfeit the re-usability advantage that the multi-model infrastructure provides. In addition, it might not be possible, feasible or desirable to convert all models to one overarching resolution. There are existing multi-model infrastructures that do go this route, for example multi-models based on high-level architecture (HLA) (Dahmann et al., 1997). One key difference between the HLA multi-model and the multi-model proposed in Multi-model.nl (2021) is that HLA's are comprised of models specifically designed for multi-modelling. This is not the case in the energy multi-model, making an overarching resolution architecture much harder to achieve. Nevertheless, coupling efforts that use HLA or full-model resolution conversion are still interesting to explore, because issues present in their resolution conversion effort might also present themselves when creating conversion algorithms.

2.2.2. Coupling differing models

All multi-models aren't created equal. The level of coupling tightness between two models can vary greatly. Brandmeyer and Karimi (2000) differentiates between five levels of coupling, which can be seen in figure 2.1 from their 2000's paper.

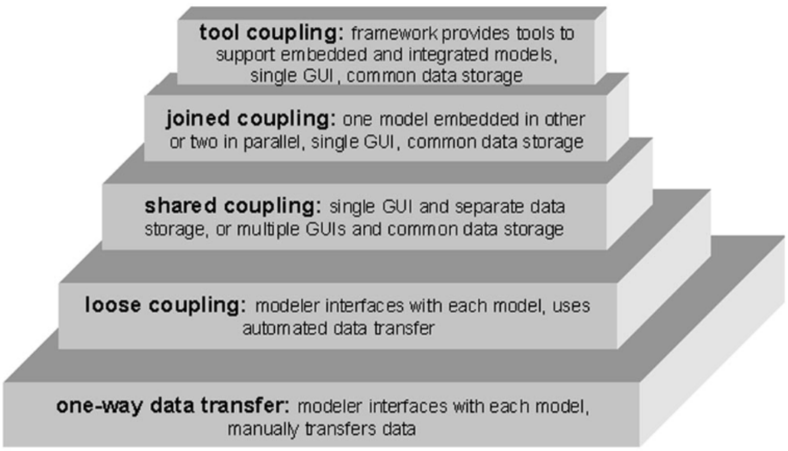

Figure 2.1: levels of coupling tightness, from (Brandmeyer & Karimi, 2000)

Figure 2.1 shows a progression in coupling tightness from the bottom tier, being one-way data transfer, to the top tier of a fully integrated tool coupling. The further up the tier the multi-model coupling is, the more integrated and shared the models and their data becomes. Whereas one could see the top tier as simply the best, the nature of the multi-modelling infrastructure (Multi-model.nl, 2021) favours a more loosely coupled approach (tier 2). A loose coupling has the advantage of retaining a level of privacy for owners of the used data and models, while still enabling an automated conversion and transfer of data between models. The convenience of higher tiers of coupling comes at the price of having to streamline the data streams and (future progression of) models between all parties involved. A feat that may or may not be feasible.

A loose coupling between two models entails that the models are coupled sufficiently tightly that automated two-way data transferring is possible during a simulation (Brandmeyer & Karimi, 2000).

This is an upgrade from the first tier, where there only exists a one-way coupling that enables manual transfer of data before or after simulation runs. A key facet to keep in mind for both the first and the second tier is that the models in question are developed completely independently and they remain independent. The coupling facilitates transfer of data between them, but neither models ”knows” that the other exists. Due to the lack of alignment in development, it can be safer to assume that there exists little to no conceptual overlap between the models. When coupling such independent systems, several issues can manifest (Tolk, 2003) such as:

- Semantic conflicts, where what is meant by a parameter is not exactly equal for both models. This could either be a nuanced difference in a parameters' meaning or unit.
- Heterogeneous conflicts, where different meta-modelling languages are used to describe the functioning of sections of the model, making comparison more difficult.

The aforementioned conflicts need to be resolved before a coupling can be made successfully. The focus and work done by resolution transformation functions can be put aside for one moment, because the issues presented here lie within the models themselves. If two models communicate with each other through the in- and output of ”energy”, can it be assumed that both models conceptually agree on the word ”energy”? Energy can be conceptualised in a myriad of ways: electrical, heat, gravitational, potential, unspecified etc. Subtle conceptual inconsistencies can make a coupling illogical without them being properly recognised and addressed. However, by no means is having differing conceptualisations a bad thing. Especially in socio-technical systems, having multiple conceptual interpretations of a system at large can aid in creating a broader multidisciplinary understanding (Seck & Honig, 2012). The point is that for this benefit to be realised, the differences in concept need to be understood first.

2.2.3. Challenges in multi-modelling

A significant area of challenge for multi-modelling is interaction between differing model resolutions. Due to the novel use of multi-modelling as a tool for the (Dutch) energy infrastructure, the shape and size of resolution-based challenges in this context remain unknown. Extensive research on proper resolution-based challenge handling has been done (Natrajan et al., 1997; Tan et al., 2001; Tolk, 2003; Yilmaz et al., 2007). One key difference is that in these applications of multi-modelling, such as in military simulations, the models used were purpose-built to interoperate (Dahmann et al., 1997). Coupling existing energy models entails using and coupling models that were never meant to interoperate. Whether this design goal difference changes the size of resolution-related problems is unknown.

2.3. Multi-resolution modelling

In modelling, resolution refers to the level of detail that is given to any (part of a) model. High resolution refers to a model that operates on a relatively high detail level, whereas low resolution operates on a more broad level. This detail level and the accompanying relative resolution of a model can be related to the total amount of parameters, inputs and outputs used (Davis & Bigelow, 1998a). When comparing two models in a multi-model for instance, the model with the higher amount of parameters can broadly be viewed as being of a higher resolution.

In addition to the resolution of any singular model, Multi-resolution models (MRM's) refer to a collection of models that operate consistently on different resolutions in order to solve a problem (Davis & Bigelow, 1998b). Multi-models, especially when sourcing models from different sources or companies, will inevitably use models with differing resolutions.

MRM's can be very useful in situations where a certain level of detail (or a lack thereof) is needed for a smooth, fast or sufficient model run (Salome, 2021), or in situations where the complete view of a problem cannot be captured by one model (Bankes, 2002). Varying resolution beforehand to fit a certain purpose can therefore improve overall MRM functionality (Natrajan et al., 1997; Rabelo et al., 2016). However it is uncertain what the effects are of combining various models with resolutions not purposefully built for MRM's in an MRM.

2.3.1. Resolution and scaling

An important distinction needs to be made between the concepts of ”Resolution” and ”Scaling”. Difference in resolution explicitly refers to two or more models describing the *same system size* on different levels of detail. For example, three models describing the energy in Netherlands: one national gas transportation model on high detail, one national household gas usage model on medium detail and one national policy model on low detail. The key is that the level of detail varies while the size of the studied system remains the same. Difference in scaling refers to the opposite: the size of the studied system varies. For example, three models describing energy: one model of a gas-fired power plant, one model describing household gas use in The Hague and one national gas transportation model.

Coupling models of varying scales is a distinctly different challenge than coupling models of varying resolutions. The focus of this paper is resolution. Therefore models depicting the same system size but different topics and resolutions will be studied. However, picking one resolution, e.g. national, does not explicitly mean that none of the lessons learned can be transferred to multi-resolution models on another scale.

2.3.2. Dimensions of resolution

The resolution of a model can vary along a series of dimensions (Rabelo et al., 2016) such as:

- Temporal: The (difference in) size of the (smallest) timestep (seconds, weeks etc.) can influence model results.
- Spatial: similarly to spatial resolution, concerning the scale of maps. Is the smallest distance interval a millimeter or a kilometer? This influences the spatial distribution of attributes in the model.
- Object: Some objects in models that are valued as more important can be given a higher resolution by means of more parameters or complex functions describing the behaviour.

The three dimensions highlighted above concern different parts of models. Each dimension has different important factors regarding the effect of resolution on model performance and results. Because of these differences, each dimension also has its own way of possibly causing issues. In order to delve into how each of these dimensions of resolution manifest and affect how models function in their own way, all three dimensions will be discussed separately below.

2.3.3. Time

Time is handled in discrete intervals in all computer-based simulations. In models such as ABM's, SD and optimization models, the total simulated time (timescale) is broken up into discrete homogeneous timesteps. Such a timestep, as described by Nance (1981) as an instant, is the smallest time interval in which a value can be assigned or changed in the model. A model cannot execute an activity (defined as well by Nance (1981)) that would take less time than the timestep size. This makes the selection of the timestep a vital decision in model creation. One way to decide on the timestep would be to look at the duration of the shortest activity in the representation of the real system and equate the timestep to that duration.

Given model spanning a simulated run of one year with a shortest activity in hours, the timestep might for example be in hours. Increasing resolution in the time dimension would be to have the timestep become smaller, lets say minutes. Reducing the instant interval to minutes has no effect on how the model is able to handle the activity. While increasing the time resolution doesn't seem to change behaviour, it can significantly decrease model run performance. However, increasing time resolution where possible might reveal interesting behaviour in certain model such as time-sensitive oscillations. On the other hand, if the resolution in time would be decreased to an instant being 2 hours, the model will now not be able to execute activities that are shorter than 2 hours. Due to the smallest timestep now being two hours, the smallest increment of change in the model will be two hours as well. The one hour activity can then only be processed as an activity that starts at $t = 1$ and ends at $t = 2$, with two hours having past in simulated time. This change in time resolution therefore changes the simulation behaviour.

Changing the resolution in time can have implications on what type of activities the model is able to do. A model running on a certain timestep cannot execute activities shorter than or in between those timesteps, not is it able to send or receive data from other models. Time is also computationally expensive. Increasing the resolution of time from being an hour to being a minute results in sixty times more timesteps having to be calculated to span the same range of time. Therefore an appropriate time resolution remains heavily dependent on the use case of the model and the timescale of activities in the real system.

Time is viewed upon and used in a variety of ways in modelling. Overstreet and Nance (1986) describes the worldview ”event scheduling” as the locality of time. Models adhering to this worldview are governed by an event calendar that schedules events and is scheduled upon by new events. Two special cases of event scheduling are System Dynamics (SD) and Discrete Time System Specification (DTSS) models. SD models pretend to operate on the basis of continuous (Euler) differential equations, while in fact an event calendar with an incredibly small timestep (Δh) between which nothing can happen lies at its core. The solver of the differential equations is often sensitive to the change in timestep, to the point where for instance oscillatory behaviour can cease to occur if the timestep were made too large. Since differential equations are also not made for sudden large disruptions, the abrupt change in input caused by a course or aggregate used dataset could impact the models' performance. DTSS models are fashioned after the principles of activity scanning (locality of state, (Overstreet & Nance, 1986)). However, DTSS models such as agent-based models in practice function along a similar event calendar, which is divided beforehand in discrete ticks. Nothing is able to happen between these discrete intervals, in contrast to for example DEVS where an event can be scheduled on any point in time. Providing a disaggregate dataset that functions on a higher time resolution than a DTSS models' ticks is therefore of no use, because the model won't be able to use any of the data between the ticks. Whether aggregation efforts in the time dimension impact DTSS disproportionately compared to other model types is unknown.

2.3.4. Space

The space dimension concerns itself on the geographical representation of distances and movements within a model. Given that space and movement play a role in the studied model, the level and resolution that is representing space can have a great impact on how the model performs and how it is able to communicate its spatial data to other models.

Spatial resolution manifests in how models handle any things from terrain maps to wind direction maps to load distribution maps. ”Map” in this instance refers to the coupling of at least one geometric and one semantic attribute (Stell & Worboys, 1998). Either or both of these attributes is able to vary in resolution. For example: a model simulates the spatial distribution of load generated from electric vehicles (EVs) charging in the Netherlands. The geometric attribute is how the model shows the spatial load distribution and the semantic attribute is the load itself. Geometric resolution dictates how granular space is represented. The Netherlands can be divided into chunks of one square kilometer, one acre, one square meter. On the other hand the modeller could choose to divide the Netherlands not by a grid but by man-made borders such as provinces, municipalities or smaller.

Assuming that a model using a spatial component is somehow dependent on spacial distribution for its results, the choice of spatial resolution influences results as well. In general, root means squared error (RMSE) increases and accuracy to source data decreases linearly as the resolution of the map decreases (Degbelo & Kuhn, 2018). However, spatial resolution can also directly influence model results. Using the EV example again, cars are expected to drive from a start point to a destination. Lets say that in this model space is displayed as one centralised node per municipality. All spatial detail within municipalities is therefore lost. Cars can only drive from one centroid to another instead of real-life destinations. This aggregation influences total distance travelled, therefore changing battery depletion and changing the main model output: charging load. In the EV model the total load is also outputted as one aggregated national load. Keep in mind that this output becomes an aggregation of an aggregation due to the previous spatial resolution choices.

Space is an important factor in models adhering to the locality of state (Overstreet & Nance, 1986). Models that deploy a practical version of this locality are ABMs. Agents in agent based models scan their surrounding environment every tick in order to determine which actions to take. Space plays a key role in this, because many ABMs have their agents scan not the entire space of the model, but a radius of e.g. patches around them. The resolution of the space in which the agents are present therefore directly affects the information that agents are able to base their behaviour on. Increased spatial resolution in a model can cause changes in behaviour by causing greater heterogeneity in the agent-scanned environment. Similarly, increased homogeneity resulting from decreased resolution decreases information variety for the agent. How exactly a change in spatial resolution translates into change in behaviour depends on how the agent is coded, which varies from model to model. Nevertheless it is important to realise the impact that spatial resolution could have on DTSS and other models.

2.3.5. Object

The object dimension refers to the level of detail that can be given to model any kind of entity. Once again, choosing a resolution depends on the use of the model. A gas-fired power plant can be modelled as a single node with an output, or as an incredibly intricate process of pressures and thermodynamics. Aggregation and Disaggregation (A/D) of objects is very difficult, because this requires changing functions themselves and writing an adjacent model at a different resolution (Salome, 2021). There is a strong overlap between the aggregation and disaggregation of entire objects and structural uncertainty. Structural uncertainty concerns how different models can represent a system in a slightly different way, which can impact model results (Draper, 1995). For non-linear models of (unintended) varying resolutions, the model states will only cohere for a limited time space and on a limited detail level (Bankes, 1993). Aggregating and disaggregating objects intentionally varies the resolution of model representations in order to function, making them fundamentally structurally uncertain. It is therefore very important to understand under which circumstances an object will need to change resolutions and for what time span.

Process interaction models define a chain of states in time that an object is able to go through throughout the model. They adhere to the locality of object (Overstreet & Nance, 1986), meaning that processes are centered around the complete series of transformations that an entity goes through. In essence it can be seen as a large chain, or set of interwoven chains, that progress slowly link by link. Altering resolution in the object dimension could affect this type of model drastically. The chain of processes could be summarized, leading to a greater state change at once. However, processes that would interact in between these state changes normally would now only be able to do so after the aggregated change have all been made by the process.

2.3.6. Resolution, why higher is not always better

When examining different models, the variety in resolution choice might come as a surprise. Why don't all models aim for a higher resolution? After all, increasing the resolution of a model can enable a closer resemblance to the real world phenomena (Davis & Hillestad, 1993). Wouldn't this always be better? Choosing an appropriate resolution for a model has several factors that influence what is ”best”. Examining these reasons provides insight on why models in an MRM vary in resolution in the first place. It could also provide insight into how models at certain resolutions could best be deployed in an MRM.

A model operating at a high resolution is aimed at resembling ”reality” as detailed as is feasible for the modeller (Davis & Hillestad, 1993). Therefore high-resolution models are able to use very detailed data as input. An example of this would be a digital twin of a gas power plant. Directly measurable details such as internal pressures, temperatures, oscillation speeds, fuel composition and airflow could all be inserted directly into the model to monitor the functioning of the plant. A lower-resolution model that simulates the power plant as a single node with a mean fuel input and accompanying output would be unable to use these detailed parameters for its calculations. However, this does by no means state that one is better than the other. It is the purpose of the model that decides what is a proper choice of resolution (Chen & Li, 2021). For example, would you need a complex description of all the pressures in a gas plant if you were only interested in its mean yearly output? The choice of resolution is something that needs to fit its purpose, which in this case favours a lower resolution.

Low-resolution modelling has several distinct advantages. First of all, low-resolution models are much more easy to oversee (Chen & Li, 2021). Easy of oversight could prove vital when the models' purpose is to be used by policy makers who are less well versed in handling complex models. Additionally, choosing a lower-resolution could reduce computational cost where possible. When using many models in a federation, such as a military simulation, the saved computational cost could be the difference between a successful and unsuccessful model run (Xuefei et al., 2017).

Aside from the fitness for purpose, resolution choice can also be largely driven by the (availability of) data. The ”proper” resolution choice for (spatial) models can be seen as dependent on how the data was initially collected (Degbelo & Kuhn, 2018). When data, like wind speeds, is natively collected per square meter per minute, modelling with wind on these resolutions would be most appropriate for the data. In addition, if a rich source of data on provincial level exists, but data on city level is sparse, a modeller could be swayed to choosing provinces as their choice of spatial resolution.

2.4. Aggregation and disaggregation

This section will discuss the various implementations found in literature of aggregation and disaggregation (A/D) functions. Many models already use some sort of A/D functions in order to present their outputs in the originally desired format. Other A/D functions are used to bridge the gap between resolutions that may occur during multi-modelling. Standard functions will be introduced that are able to function well for basic A/D. attention will also be given to how A/D can be tailored specifically to a model, in order to produce a more realistic A/D effort.

2.4.1. Aggregation and disaggregation fundamentals

Aggregation and disaggregation (A/D) functions act as the transformations medium through which two models of varying resolutions communicate. In essence, A/D functions can be quite straight forward. Usually a function such as a mean, median, max, min or distribution is used to coarsen or fill in the original data, thereby changing its resolution to the level that the input of another model can receive (IBM, 2021). In order to increase the resolution, data points are often split according to a distribution centering around a mean. A simple example of such a function would be the aggregation and disaggregation of a sensor measuring temperature each minute for 15 minutes. In order to reduce the resolution of the data to one 15-minute interval one could take the mean of the batch of fifteen 1-minute measurements like in equation 2.1. Then, to increase the resolution in a simple fashion, one could disaggregate by labeling the mean as being every value for those fifteen 1-minute intervals like in equation 2.2. Probability distributions such as the normal distribution are also a popular method to disaggregate data with.

$$T_{agg} = \sum_{i=1}^{N} \frac{T_{i_{minute}}}{N} \tag{2.1}$$

$$T_{disagg} = T_{agg} \tag{2.2}$$

Basic operations are commonly used and often adequate to bridge the gap between resolutions in multi-resolution simulations (Xuefei et al., 2017). However, it cannot be assumed that all manner of parameters can simply be converted using a transformation such as a mean. Many A/D functions don't incorporate the origin of the data or even the desired end resolution in favour of using a 'general' A/D approach (Skogan, 2001). Skogan (2001) argues that the origin and datatype of the input as well as the datatype of the desired output need to specifically taken into account to tailor an A/D function to the data it will be used on. It cannot be assumed that because a list of data is of the same type and length as another list, it can therefore be aggregated and disaggregated in the same way instantly. For example, a model simulating power demand growth D per year for 10 years might produce a list of data similar to the input of equation 2.1. The sample data is presented in 2.3.

$$D_{growth} = [1.05, 1.1, 1.42, 1.3, 0.89, 1.12, 1.08, 0.81, 0.94, 1.24] \tag{2.3}$$

The data in 2.3 is of a shape and type that is able to be handled by the standard A/D functions of 2.1 and 2.2. If equation 2.1 is applied it would successfully yield an aggregated mean growth of $D_{agg,\ mean} = 1.095$. However, the result would be inconsistent. Over the full timespan, disaggregation by equation 2.2 would set the mean as every value. This is inconsistent with the actual mean growth as seen in equation 2.4 and 2.5.

$$D_{agg,\ prod} = \prod_{i=1}^{N} D_{growth_i} = 2.167 \tag{2.4}$$

$$D_{disagg} = \sqrt[N]{\prod_{i=1}^{N} D_{growth_i}} = 1.080 \neq 1.095\ (D_{agg,\ mean}) \tag{2.5}$$

The fault in this approach is that the data was assumed to be linearly aggregatable purely on the datatype. Basic A/D functions have the potential to quickly transform simple sets of data. However, for non-linear equations, careful attention must be given to preserving the non-linearity wherever possible (Salome, 2021).

2.4.2. Aggregation and disaggregation constraints

Much can be achieved for pure A/D by using relatively simple mathematical functions. However, when it is desirable to more accurately perform aggregation and especially disaggregation functions, it is very important to understand the model being studied. Understanding not only the model, but also the natural system that the model is based on, can aid in creating more realistic A/D functions (Li et al., 2008). This requires a deep dive into every model that one would want to perform A/D on. To name some military examples, frameworks of functions for aggregation and disaggregation have been specifically documented for aircraft models (Li et al., 2008) and armoured vehicles (He et al., 2014). A key concept from these papers is the notion of constraints. Li et al. (2008) distinguishes between two types of constraints:

1. System specific constraints
2. Logic specific constraints

System specific constraints

System specific constraints are methods that ensure that A/D efforts don't create a bias away from the original data based on their mathematical construction. An excellent example is the *equal average rule*, which states that the average of the disaggregate and aggregate values must be consistent. When this rule is not upheld like in equation 2.5, rapidly oscillating A/D, called thrashing (Chua & Low, 2009), would distort the flow of information between models severely. Thrashing can occur when a model itself switches resolution, but also when a set of models frequently pass along a data while having to switch its resolution for every pass. When the equal average rule *is* upheld, A/D thrashing can still rapidly change the data back and forth, but it won't be able to structurally drift away from the data that the A/D functions revolve around. Take for example equation 2.6. The equation describes the position of a group of vehicles, and states that the aggregate position must always be equal to the average of the disaggregate positions.

$$\frac{1}{N}\sum_{i=1}^{N} Positions_{E_i} = Positions_{agg_E} \tag{2.6}$$

An accompanying disaggregation function can take this aggregate position and create a disaggregate set of randomized positions for every vehicle, with the constraint of having the mean of those new positions still equal the same mean. If thrashing were to occur in this situation, the positions of individual vehicles would still change every A/D cycle as if they were teleporting around. Is this an issue however? A/D usually occurs because of a trigger event that sets off the transformation (Chen & Li, 2021). In this case, the vehicles were to aggregate whenever they are far away from a battle and disaggregate when near enemies to simulate combat. Rapid A/D can be interpreted as an enemy losing sight of the squadron of vehicles and therefore they could be in a range of superpositions. All these possible disaggregate location configurations are valid according to the equal average rule. Once the 'enemy' comes into a certain close range, the final disaggregation transformation would create the final position set of the vehicles. Disaggregation through a random drawing process can be considered consistent within this system.

System specific constraints lay the foundation of a consistent A/D method (Chen & Li, 2021). This is why almost all A/D functions adhere to system specific constraints (He et al., 2014; Li et al., 2008; Reynolds & Natrajan, 1997; Xuefei et al., 2017). In some cases, constraints such as the equal average rule cannot be upheld. A/D can then still be done with a threshold in place to determine what is an acceptable level of A/D drift (Gou et al., 2015).

Logic specific constraints
Logic specific constraints, in contrast to system specific constraints, are designed to aid in increasing the realism of A/D. Adding logic specific constraints to disaggregation functions in particular has the potential to fill up the gap of the information lost while aggregating with more realistic disaggregation results, without having to store more information during a run. For example, the speed v and trajectory of a squadron of N individual aircraft can be aggregated to one aggregated speed using equation 2.7 (Li et al., 2008).

$$v_{agg} = \sqrt{\left(\sum_{i=1}^{n} v_i * cos(\alpha_i)\right)^2 + \left(\sum_{i=1}^{n} v_i * sin(\alpha_i)\right)^2} \tag{2.7}$$

Li et al. (2008) adds that a logic specific disaggregation effort to accompany equation 2.7 can increase realism of results without forfeiting system specific constraints such as the equal average rule. For example, any aircraft type has a maximum and minimum flying speed (U.S. Air Force, 2015). Instead of disaggregation by completely random drawing with a mean average rule, data such as from the U.S. Air Force (2015) can be added to create for example a normal distribution hovering around the aggregate mean, with any values above the max and below the min speeds removed. Similarly, a positional disaggregation accompanying equation 2.6 can be enhanced by knowing that any aircraft in formation must keep a certain distance from other aircraft, while staying within a certain vicinity of the squadron leader (Li et al., 2008).

Knowing the logic behind the entities that A/D is done on can aid in creating more accurate A/D functions. However, a trade-off has to be made between the increased realism and the increased static consistency cost (see section 2.5.3) of creating these more detailed functions.

2.5. Consistency

For two models of differing resolutions to communicate, A/D transformations will need to be done to align the inputs and the outputs to each other. A goal to strive for is for these transformations to be ”consistent”. Consistency is often mentioned in military research, where entire models are transformed to another resolution, instead of just the output parameters. These models are deemed 'completely' consistent if the *end states of both the low-resolution and high-resolution iterations of the model produce the exact same output results when compared on either of these resolutions* (Davis & Hillestad, 1993; Reynolds et al., 1997). For energy multi-modelling the definition can be slightly rephrased. 'Complete' consistency in models that only aggregate and disaggregate outputs can be seen as that a subsequent aggregation and disaggregation (A/D) effort done on an certain output would yield exactly the same set of values as before the transformation was done. An example of such consistency was already provided in 2.4.1.

Given the sparsity of literature on energy multi-modelling, this section will examine the available literature on how consistency issues were addressed in other fields such as military simulation on which extensive literature exists (Natrajan et al., 1997; Tan et al., 2001; Tolk, 2003; Yilmaz et al., 2007). Relevant lessons learned from these papers can then be drawn to aid in creating consistent couplings for energy multi-models.

2.5.1. What is proper consistency?

Consistency between aggregated and disaggregated models is often categorized into strong or weak consistency, where the former suggest a better resemblance between different resolution tiers of models and the latter potentially enabling a faster throughput (Adya, 1999). Strong consistency is often seen as the goal to strive to. However, this assumes that all source data is complete and high-resolution at the start, which is not always the case. In multi-resolution multi-models, information is inputted, processed and outputted from all different kinds of resolution levels. For example in an energy system, a broad 'low resolution' political input to reduce emissions by 50% in 10 years cannot be treated as more or less important as a 'high-resolution' input such as current incoming power from a wind turbine. The nature of MRM's is that information is handled at different resolutions. Therefore, placing ‘high-resolution’ information above ‘low-resolution’ information as a judge of consistency is inadvisable (Davis & Bigelow, 1998a).

Reynolds et al. (1997) states that, whether strong or weak, complete consistency should be strived for, but not how an model receiving and sending information across multiple resolutions can be kept consistent (Yilmaz et al., 2007). But what is meant by ‘consistency’ across resolutions? An intuitive answer would be that subsequent A/D would return the same model or output state, thus making the process fully reversible as proposed by Reynolds et al. (1997). Ideal as it would be, it is deemed unfeasible and some research even argues that it is unneeded (Davis & Bigelow, 1998a). Davis and Bigelow argue that ”good” consistency should not be judged on whether they generate exactly the same system state, but on whether they generate the same behaviour when interacting with other models in the multi-model. ‘Consistent’ then becomes a much more subjective term. Disaggregation and aggregation might not be perfectly reversible, but that is a non-issue if it has an understood and acceptable influence on multi-model behaviour.

A good pointer to base these decisions on could be sensitivity analysis. Sensitivity analysis tools are able to examine which inputs (or change in those inputs) create the most significant changes in outputs (Zhang et al., 2015). If the input receiving data through A/D is able to create large changes in output through small changes in its input, increased attention to a proper and more tightly consistent A/D could be advisable. However, one could also turn the argument around and say that sensitivity of the coupled inputs is a requirement to start coupling in the first place. A key argument for using multi-models at all is that multi-models can couple existing models which can all contribute meaningfully to each other to create a more integral view on a complex system (Multi-model.nl, 2021; Seck & Honig, 2012). The question is then if a coupling is even the right thing to do when the inputs to be coupled are (relatively) insensitive to what they’re receiving from the other model. Yes, insensitive parameters could possibly be coupled using courser means of A/D without causing consistency issues. But with this being said one could wonder, having encountered such insensitive coupling inputs, if these are the right inputs or even the right models to be coupled.

2.5.2. Consistency issues in aggregating and disaggregating models

A critical factor in combining MRM’s lies in inter-model communication. In order for models of differing resolutions to communicate, models will have to adapt to match each other. However, running high-resolution models only to then aggregate the results is quite inefficient. Instead, one complete aggregation or disaggregation of the model would be more efficient (Davis & Bigelow, 1998a). This means creating multiple versions of the same model on different resolution levels. An example from military modelling would be a camp of soldiers modelled on two resolutions:

- Low resolution, modelling only with the camp’s coordinates and population count
- High resolution, including layout, deployment location of troops, present ammunition etc.

A similar use case can be made for energy models at different resolutions. For example, when simulating a fleet of electric vehicles and monitoring grid loading behaviour.

- Low resolution, modelling only one national electric vehicle loading curve in time
- Mid resolution, modelling electric vehicle loading curves per municipality in time
- High resolution, modelling each individual vehicle with its accompanying loading behaviour

Once multiple resolutions depicting the same model have been created, the question arises when (and how) to switch between these resolutions. A method used in the past was Full Disaggregation, in which any and all low-resolution models (LRM’s) disaggregate to a higher resolution when interacting with a high resolution model (HRM) (Reynolds et al., 1997). However, Full Disaggregation has the downside of potentially causing a chain disaggregation, due to all models that previously interacted with the LRM now also having to disaggregate. On top of that, switching often between resolutions could create unrealistic system states (Natrajan et al., 1997). As depicted in figure 2.2, small inconsistencies in A/D can introduce ever-increasing bias as the model repeatedly shifts between resolutions. Very frequent A/D, also called ”thrashing”, can further enlarge possible inconsistencies (Chua & Low, 2009).

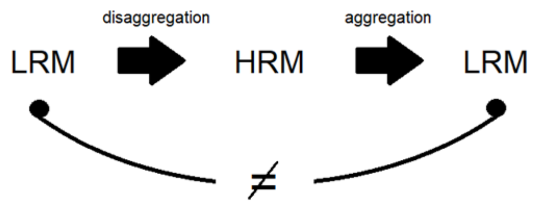

Figure 2.2: hazard of repeated aggregation and disaggregation

In order to counter inconsistency issues generated by the Full Disaggregation technique, Reynolds et al. (1997) proposes creating a Multi-Resolution Entity (MRE) instead. An MRE is a single, internally consistent entity comprised of multiple levels of resolution all representing the same system. Instead of switching from LRM to HRM when needed, both the LRM and HRM exist inside the MRE. Models of differing resolutions can interact with the appropriate resolution of the MRE. In the MRE concept, a consistency enforcer inside the MRE ensures that the state of the MRE is made consistent across all levels of resolution after every interaction. This consistency enforcer uses A/D mapping functions that are tailored very specifically to the system in question. These A/D mapping functions can also be used to communicate outputs between different models.

Using an MRE has the advantage of isolating the component responsible for internal consistency, namely the consistency enforcer (Natrajan et al., 1997). This has benefits over previously used ad-hoc mapping approaches, as such an enforcer could be designed with a somewhat general and modular thought in mind. Natrajan proposes using an attribute dependency graph to couple all resolutions together. A set of attribute nodes for each node is generated and connected by weighted nodes that describe the transformations needed. During a run, these mapping functions can then be used to carry information from one resolution to the next (Jie et al., 2012). Ideally, these A/D transformations are reversible (Reynolds et al., 1997). Ensuring consistency in an MRE requires a very tight coupling through the consistency enforcer. A disadvantage of this method is that MRE extension after the fact is difficult (Yilmaz et al., 2007). Therefore, a thorough understanding of the needed resolutions is necessary before any MRE or A/D is created.

2.5.3. Consistency cost

Following the proposed attribute dependency graph by Natrajan et al. (1997), communication and consistency maintenance consists of mapping functions and aggregation/disaggregation of data. These functions impact model performance in two ways: Static Consistency Cost and Dynamic Consistency Cost (Natrajan et al., 1997). Static consistency cost regards the initial time cost that is needed to create the mapping functions. It draws from the observation that a tight and consistent coupling is time-consuming to create and expand (Yilmaz et al., 2007). Figure 2.3 shows how mapping functions can be designed. The left of figure 2.3 shows mapping functions between single resolution levels, creating a ladder through which data can be exchanged. The right of figure 2.3 shows a full graph approach, where the MRE is able to 'jump' from one resolution to another without going through intermediate resolutions.

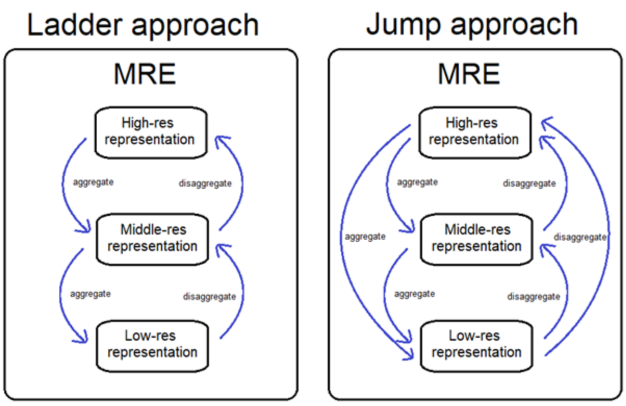

Figure 2.3: internal MRE mapping approaches

Dynamic consistency cost regards the time cost of using the mapping functions to keep the MRE consistent during a simulation run. Especially with MRE's with many resolutions, the 'jump' approach of figure 2.3 is potentially faster, resulting in a lower dynamic consistency cost combined with a higher initial static consistency cost due to increased build time. Models that communicate often stand to benefit additionally from these jumps, as the added dynamic cost of thrashing is reduced. However, the jumps approach introduces a significant risk of overall consistency. Using the jumps approach there are now multiple mapping paths to aggregate and disaggregate from one level to another. Assuming that these paths cannot always create perfectly similar results, the 'jumps' method introduces inter-resolution mapping inconsistency errors for the sake of simulation speed. Especially in multi-models with rapid communication and thrashing between resolutions, the gain in model performance might come at the cost of the accuracy of model results. This trade-off between consistency and speed goes against the initial concept of the MRE to be as reversible and consistent between resolutions as possible (Reynolds et al., 1997).

Furthermore, dynamic consistency cost for the coupling of stochastic models may increase. Stochastic models are able to vary based on the same inputs and a pseudo-random generator. When coupling a stochastic model, one might want to compensate for any outliers in the specific instance that the stochastic model is running. If outliers are a significant worry for a coupling, compensation through additional runs and consistency checks could increase dynamic consistency cost.

2.5.4. Consistency checking

Multi-resolution models often employ consistency checking methods such as in figure 2.4 (Gou et al., 2015). In a multi-resolution model such as an MRE, model A and model B would be two instances of the same model operating at different levels of resolution. The output similarity would then have to fall within an accepted margin of error to be considered sufficiently consistent.

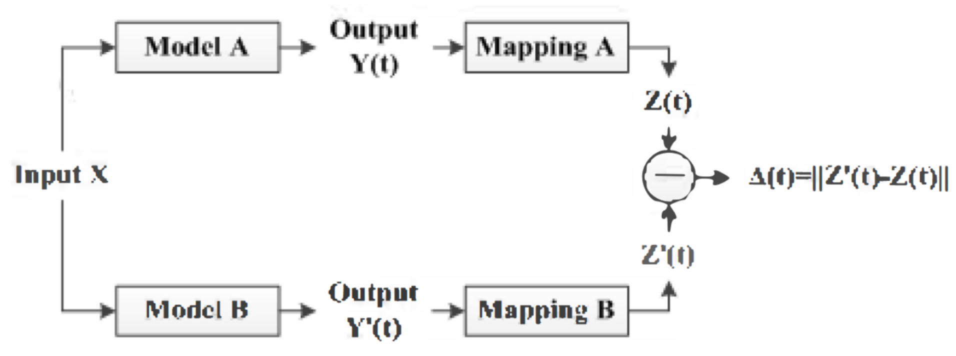

Figure 2.4: Output comparison of two models, by Gou et al. (2015)

Multi-resolution models compare the outputs of two instances of the same model on different resolution. In multi-resolution multi-modelling, this is impossible since each model only has one level of resolution. However, the comparison method as described in figure 2.4 is widely used in different multi-resolution consistency checking problems (Chen & Li, 2021; Davis & Hillestad, 1993). Based on the consistency checking frameworks described in literature, an adaptation could be made that would function in the use-case of multi-resolution multi-modelling. Figure 2.5 displays such an adaptation. Given a model sending out information on a low resolution which needs to be converted to a higher resolution, the consistency of the disaggregation effort can be tested in this way. After the disaggregation effort is done, the disaggregated data A' can be re-aggregated to A" which enables a direct comparison to the original data A. If A and A" differ largely (beyond the set error threshold E_{thresh}), this could indicate that the disaggregation effort is affecting the data in an undesired way.

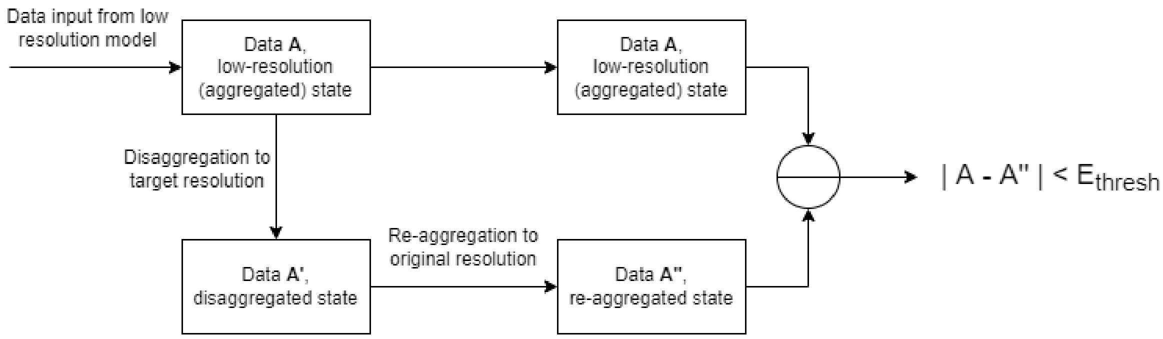

Figure 2.5: adaptation of output consistence method

2.6. Literature findings

Multi-resolution multi-modelling has been covered quite extensively in literature. Primarily, it becomes apparent from the readings that a strict manner of modelling is used in applied multi-resolution multi-models. Strict order in terms of object and semantic definitions reduce consistency-based errors in these models. When it comes to energy modelling however, literature is sparse. The challenge with energy multi-modelling seems to revolve around situations where these strict modelling principles were not upheld beforehand. Such challenges are especially present in the models used in the multi-model infrastructure project, as each used model was created completely independently.

3

Method and case selection

Now that a literature study has been performed, the next step is to apply the findings and principles on a suitable case of multi-resolution energy model coupling. This chapter contains the methods used to gain the information needed from pre-existing models to be able to pinpoint possible problem areas in coupling. In addition, a case study of energy models is selected to test the methods on. This case study aims to identify possible additional coupling problems, and to create a functional coupling that relieves these issues where possible. The effectiveness of these methods will be discussed in the subsequent chapter.

Regarding the coupling method, literature shows the level of detail and understanding required to properly aggregate a model (He et al., 2014; Li et al., 2008; Reynolds et al., 1997; Salome, 2021). From this it becomes more and more apparent that each model and each coupling will require a mapping that is specifically created on a case by case basis. What could aid in generalizing this process is to document the steps of reasoning used. Even if applied to different cases, a more homogeneous approach to model coupling could serve as a first step in generalizing the coupling itself.

3.1. Model specific auditing

Mao et al. (2013) states that the key advantage of multi-resolution multi-modelling is that given a certain situation, you are able to choose the appropriate resolution to fit the task. HLA is then used to align semantic, syntactic and pragmatic choices over a large span of models (Jie et al., 2012; Tan et al., 2001). Since the models presented for the multi-model infrastructure lack this alignment, certain facets of a models' working will have to be identified prior to coupling. Only when the similarities and differences between the to be coupled models are identified, can a coupling be made.

The literature examined in chapter 2 points to a few areas of interest for possible resolution related problems, such as:

- Understanding the (differences in) the general ideas and views that lead to the creation of the models
- Syntactic, semantic and pragmatic (in)consistencies
- Differences in implementations in the resolution dimensions of time, space and object
- Consistency issues stemming from the A/D functions connecting the models

For each model to be coupled, information on these problem areas will be gathered. A case is assumed where the private party owning the model might only be partially able or willing to provide insight into a model. The model itself can then be seen as a grey or a black box model. Gathering knowledge on these types of models is often done via model 'auditing' (Adler et al., 2018). Therefore an auditing style of information acquiring is also used here. Table 3.1 shows a list of auditing questions ordered per category. The aim of the auditing questions is that for each model the results of the inquiries presented in the list creates a model 'profile'. These profiles can be held side by side. This facilitates easier comparison and identification of possible problem areas.

Table 3.1: Model specific auditing questions

Category	Audit question	Desired information	Further reference
general information	What was the initial purpose of the model?	Identify whether there is a semantic and conceptual (mis)match	Tolk (2003), Seck and Honig (2012)
	What A/D efforts have already been made to the outputs of the model?	Identify whether source data at a higher or lower level already exists. How did A/D present in the model affect output?	2.4
	On which modelling principles (and/or locality) is the model based?	Different localities have unique sensitivities to resolution changes that need to be kept in mind	Nance (1981)
	Is the model deterministic?	(How) does the model deploy randomness?	2.5.3
	What is the percentage of model in/outputs to be mapped vs total in/outputs?	Is the model going to be fully dependent on the other model or not? Coupling tightness affects static consistency cost	Yilmaz et al. (2007)
Time specific	How is time represented in the model?	Identify whether large disparities exist in the time resolution dimension.	2.3.3 , Nance (1981)
	What are the timestep, timescale and total runtime?	Identify some key values in how time is handled in the model	2.3.3
	Through which timeframe is the model expected to communicate?	Do any A/D functions need to be consistent over the whole timescale or not?	Bankes (1993)
	At what frequency is the model expected to communicate?	Very frequent communication requires increased consistency checking to prevent thrashing-related consistency loss	Chua and Low (2009)
	When is the model expected to communicate?	What are triggers for a communication to occur?	Chen and Li (2021)
Space specific	How is space represented in the model?	Identify whether large disparities exist in the space resolution dimension.	2.3.4, Stell and Worboys (1998)
	Does the model use geometric map-based data structures?	Identify whether input/output data structures in the models are geographically distributed and whether these function at a similar level of granularity or not	2.3.4
Object specific	How is the level of detail distributed across the model?	Are certain parts of the model given a higher detail level than others? Are the parts of the model that are of interest for coupling modelled in a relatively high or low detail level?	2.3.5
	How are processes that appear in both models (if any) conceptualised and/or modelled?	Does a possible resolution difference affect how certain shared processes are conceptualised and modelled?	2.3.5
Sensitivity specific	How sensitive is the model to the inputs that it needs the other model(s) outputs?	Identifying sensitivity to input parameters could aid in deciding a proper level of detail from A/D	2.5.1
	How sensitive is the model to variations in input resolution?	Coarse inputs could cause different behaviour compared to smooth inputs. If not, is a rough A/D effort sufficient?	2.5

It is important to note that the results for table 3.1 might not directly be translated into the final A/D functions. These auditing questions are also made to provide insight in what coupling these models entails by itself. For example, a given model 1 and 2 are to be coupled. Model 1 has a very strict definition of space. It functions based on geographically distributed agents interacting in a heterogeneous space with varying attributes. Model 2 is needed to provide spatial information for these attributes, but model 2 itself does not operate with a strict definition of space. It could for example function by means of a weighted graph of nodes and edges. The weights of the edges might be based on some distance in the real system, but the edges aren't truly spatially distributed. Model 1 and 2 have a conceptual mismatch that may not be resolvable. The models could still be coupled, but it is then even more important that the modellers and the users of the multi-model have the insight of the conceptual mismatch. Without clarifying these type of issues, the multi-model might be given more credit or validity than it actually deserves.

3.2. Coupling specific auditing

Once the profiles in table 3.1 are created and compared, the coupling effort can proceed to the next step: creating appropriate A/D functions. In a similar fashion, key information needed to construct A/D functions is obtained through a (coupling specific) auditing list, shown in table 3.2. Table 3.2 is more practically-oriented than 3.1. It aims at establishing specific input/output pairs and creating a 'consistent' coupling through A/D functions. This table therefore meant to be filled in once per pair of models to be coupled. In short, table 3.1 attempts to uncover more fundamental and theoretical causes of issues. The coupling specific audit in table 3.2 focuses on the more practical side of the coupling and the A/D related issues that stem from actual coupling.

In addition to the A/D functions themselves, a way to check whether the transformations done are consistent needs to be in place. Consistency checking functions, along with insight on the cost of creating (or possibly expanding) consistent A/D functions. All these functions depend on what 'consistency' means in the context of any specific coupling. Is it possible to create fully reversible A/D functions? Even if that is so, is that truly needed? As discussed in 2.5.1, a choice of appropriate consistency is not always straight forward. The choice might be based on mutual discussion and agreement between the parties owning the models. Another manner might be to examine whether creating a coarse or fine-grained aggregation significantly changes multi-model behaviour. Section 2.5.1 highlights that sensitivity analysis can be a good tool to gain insight into which inputs are most sensitive and therefore could be worth a more detailed and tightly consistent A/D effort.

Table 3.2: Coupling specific auditing questions

Category	Audit question	Desired information	Further reference
Coupling pairs overview	What are the inputs and respective units required for the coupling?	Identify which input ports will require data from another model	2.4.1
	For each input, what is the corresponding output and respective unit in the other model?	Identify which outputs will provide the corresponding data	2.4.1
	Are the semantics, syntactics and pragmatics of each input/output pair consistent?	Identify semantic, syntactic and pragmatic (mis)matches per input-output pair	2.2.2, Tolk (2003)
	What levels of resolution can the models be described as having?	Broadly identify in which direction the coupling will require aggregation or disaggregation functions to match the inputs	Davis and Bigelow (1998a)
A/D function requirements	What are the system specific A/D constraints present for each input/output pair?	Identify the mathematical manner of maintaining consistency between subsequent aggregation and disaggregation	2.4.2, Li et al. (2008)
	What are the logic specific A/D constraints present (if any) for each input/output pair?	Identify whether logic-based additions to the A/D functions exist that could increase realism	2.4.2, Li et al. (2008)
	Are there input/output pairs with constraints that are sufficiently similar that an A/D function could be re-used?	Any possible re-use of A/D functions could decrease the time cost of creating the coupling	2.5.3, Natrajan et al. (1997)
	What are the estimated input/output data sizes?	Potentially large data sizes need to be kept in mind when constructing A/D functions	2.4
	What is a realistic domain for each input and output?	Baseline check to see what the expected domain of the disaggregate will be and to check whether one model create an output that the other model is unable to process	2.4.2
Consistency maintenance	What is an adequate goal of consistency to strive for in this model coupling?	A choice has to be made regarding what level of consistency between A/D functions is satisficing	2.5.1, Reynolds et al. (1997), Davis and Bigelow (1998a)
	What is an acceptable margin of error between aggregated and disaggregated values?	Are there areas where a margin of error between A/D functions is needed? If so, is there literature that could set an acceptable error size?	2.5.4, Chen and Li (2021)
	Are the A/D functions sufficiently able to withstand thrashing?	Testing whether rapid aggregation and disaggregation causes any problems. This can be done using a dummy dataset per A/D pair	2.5.2, Chua and Low (2009)
	What are the static and dynamic consistency costs?	All A/D functions will have to be constructed and validated to examine whether they function properly. If this costs a considerable amount of time, it is useful to examine the underlying cause	2.5.3, (Natrajan et al., 1997)

3.3. Case study selection

Two separate case studies are performed in order to identify possible additional coupling problems and to create a functional coupling that relieves these issues where possible. The models in the case study were selected based on a few criteria:

- The models must vary in resolution sufficiently to provide a challenging gap to bridge
- The models must be energy-modelling related
- There must be a feasible case for why one would want to couple these models
- The models must be readily available to research

The final models selected for the case study, based on the criteria above, are listed below. Each model will be given a short introduction here, followed by an in depth section in chapter 4.

- **The energy transition model (ETM), by Quintel (2021)**
  The ETM is an exploratory model used to provide governments, companies, students and schools

a fast and user-friendly tool for gaining insight in possible future energy scenario's in the Netherlands (Quintel, 2021). It incorporates a variety of industries and sectors for the supply and demand of energy that combined provide a complete overview of energy production and expenditure, alongside statistics regarding CO2-emissions and realisation costs. The model is available online. Users are able to change future supply and demand via sliders and see the results within seconds.

- **Hydrogen-buffered Wind Power model (HWP), by Boereboom (2020)**
  The HWP is a linear optimisation model used minimize power generation imbalance cost based on small mistakes in wind-pattern predictions. The model aims to balance the imbalance between the agreed quantity of electricity sold on the day-ahead market with the actual production by means of an internal hydrogen buffer. The model is made in the modelling program Linny-R.
- **EV-power demand model (EVM), by Boendermaker et al. (2022)**
  The EVM is an agent-based model used to explore how electric vehicle (EV) power demand *and* EV battery available capacity is distributed in time and space in the Netherlands. each electric vehicle and its corresponding owner make up an EV-agent, which has unique loading and travel behaviour throughout the days. The model is made using Python and the ABM library Agentpy.

The original reports and models of the EVM and HWP are available on Github. The ETM model is available online. As a note of clarification, the HWP and EVM models have been constructed completely (in case of the HWP) or partially (in case of the EVM) by the author of this MSc thesis. The choice to re-use these models was made due to their fit to the model criteria mentioned above, specifically their availability and their relation to energy modelling.

3.4. Case studies overview

This section lays out which couplings will be made in the case studies. Each coupling has a specific use case and an accompanying exchange of information between the models. The goal of these couplings is to explore and improve the ability of the two audits to locate possible issues in multi-resolution multi-modelling. The couplings are therefore 'hypothetical' coupling scenarios. While both proposed couplings are made to be feasible, they do not serve a designated research function beyond exploring and improving the auditing method.

3.4.1. Case 1: the ETM-HWP coupling

The ETM is able to produce electricity price curves for future years, such as 2050. The HWP uses historical prices and wind speeds to derive the profitability of a windfarm. However, it is unable to predict what the electricity price or wind speeds in the future will be. The hypothetical case for the coupling of these models is therefore that an energy company would like to couple their model to the ETM in order to get price projections to estimate future profitability on. In the hypothetical case, a one-way coupling will be made that uses the ETM to supply information about future power production and prices to the HWP model. An overview of the coupling is displayed in figure 3.1.

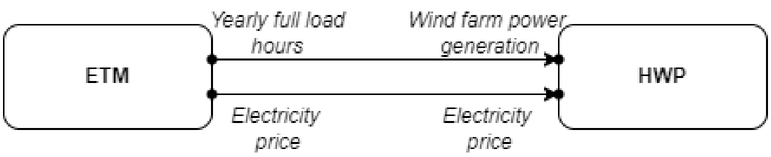

Figure 3.1: ETM-HWP coupling overview

This coupling will be relatively static, meaning that no dynamic exchange of information between the models will take place. The ETM will provide data for two different years, a starting year (e.g. 2022) and a future year (e.g. 2050). The coupling provides interesting opportunities to investigate how differences in resolution can be identified and resolved.

3.4.2. Case 2: the ETM-EVM coupling

Similar to 3.4.1, the ETM-EVM coupling will also use future scenario's of the ETM for pricing purposes. As displayed in 3.2, the ETM can supply price and population information to the EVM. The EVM is able to use that information to model future EV electricity demand and storage supply curves. Therefore the hypothetical case for coupling the ETM and the EVM is for the ETM collaborate with the EVM in order to make more detailed predictions of electricity demand and storage within the EV sector. The ETM-EVM coupling is a nice addition to the ETM-HWP coupling because of its spatial component. The EVM distributes agents across space on a municipality level, while the ETM has a national approach. The ETM-EVM coupling can therefore provide additional insight into spatial resolution-based challenges that the ETM-HWP might not be able to illustrate.

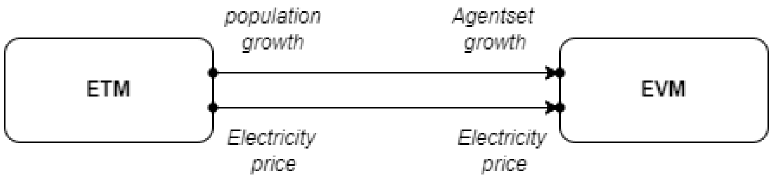

Figure 3.2: ETM-EVM coupling overview

4

Model audits

In chapter 4, the proposed audits from 3 will be implemented. The auditing of both hypothetical couplings will be documented in parallel. During this effort, the research goal of identifying and resolving resolution-based issues is ever kept in mind. The full auditing and coupling process is split into three overarching steps.

1. **Model auditing (4.1, 4.2, 4.3)**
   Each of the three models in the case study will have a full model audit. Using the questions from 3.1 as a guide, key elements relating to model resolution are investigated per model. The audits can then be compared side by side to highlight model-based differences between models that can cause issues. Finally, methods will be proposed that can alleviate the uncovered model-based resolution issues.
2. **Coupling auditing (4.4, 4.5)**
   Each model coupling pair (case 1 and 2) will undergo a coupling audit as well. Coupling auditing is already more oriented towards the practical side of the coupling. The variable-based issues of the two couplings are layed out, alongside appropriate mitigation strategies.
3. **Coupling implementation (5.1, 5.4)**
   Armed with knowledge regarding possible issues and with the means to alleviate them, the actual coupling can be created. The initial results of the coupled models runs will be displayed in this chapter. Further validation and examination is also done in chapter 5.

The steps as stated above result in three model specific audits and two coupling specific audits, which all have separate sections. At the end of each section, a summary table is added that provides a short answer to the questions asked in the audits. The model audit summaries can be found in sections 4.1.4, 4.2.6 and 4.3.6. The coupling audit summaries can be found in sections 4.4.5 and 4.5.5.

4.1. ETM model audit

The following subsections highlight facets of the ETM model by category, based on answering the questions in table 3.1.

4.1.1. General information

The ETM is an exploratory model used to provide insight in possible future energy scenario's in the Netherlands (Quintel, 2021). The model does not aim to model the viability of any one technology or sector, but to serve as an interactive exploratory tool. The model uses a plethora of sliders and hourly curves as input. From these inputs the ETM outputs some key statistics for the intended future year such as CO2-emissions, energy use etc. Most outputs available to the user are aggregated statistics as mentioned before. However, graphs representing hourly patterns for the target year are also displayed. The model is deterministic, meaning no randomness is used to create varying outputs for the same input set.

The ETM model is based on the modelling principles of network (graph) balance. The entire ETM functions as one large graph with nodes and edges representing energy use/supply and transportation (Quintel, 2022b). The user is able to edit energy demand and the availability of supply and the entire graph. The ETM then decomposes the single adjusted demand slider into smaller subsections which in turn traverse the graph from left to right until a supplier is reached. A subsection of the full graph can be seen in figure 4.1.

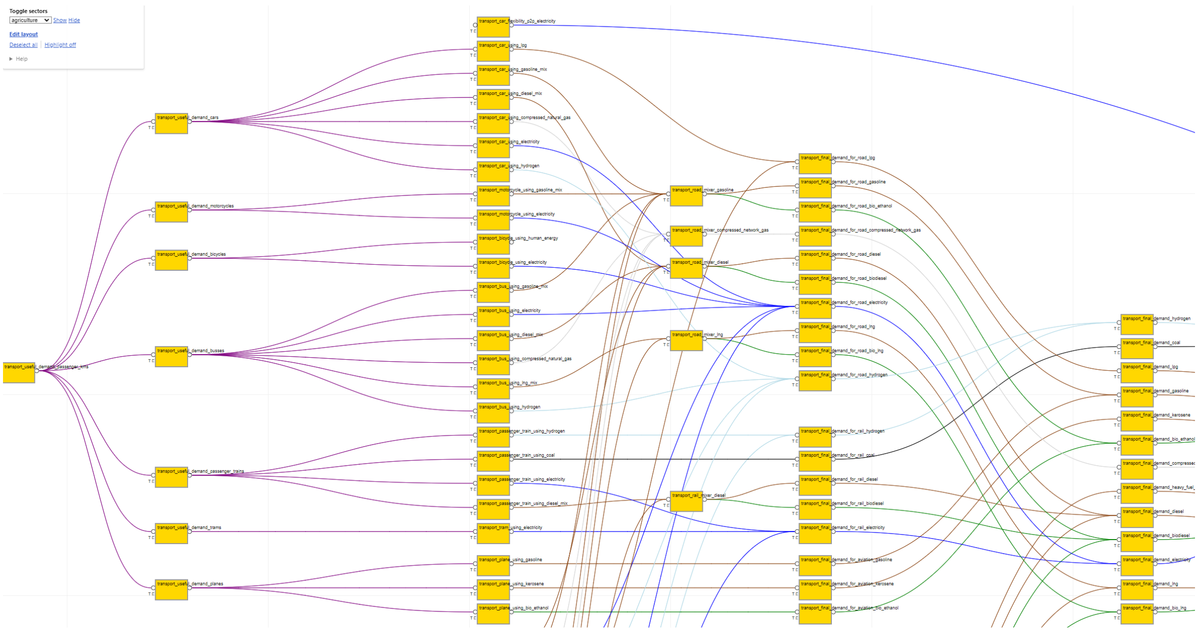

Figure 4.1: Subsection of the graph underlying the ETM (Quintel, 2022a)

Naturally, such a thorough approach requires a considerable amount of data from a range of fields and sources. The ETM uses an estimated 694 inputs, of which 62 input curves (hourly data) and 632 user-adjustable sliders (Quintel, 2022c).

4.1.2. Time specific information

The ETM's vast detail comes into conflict with its ambition to be interactive and real-time when the issue of time is represented. Time is not modelled in the ETM as it would in many other types of models. No timestep exists, nor does the model solver use iterative step-based calculations. Instead, the entire graph is solved from the start point to the end point. In other words, the set of balanced inputs and outputs at the start is altered with new inputs (for a future year) and the graph calculates the supply which would bring the graph back to balance. Therefore it is not possible to interact or stop the model in between the start year and the end year, because there is no in between.

One important note is that while not using hours as a timestep of calculation, the ETM does utilize hourly input curves in the run. These are mostly to represent daily fluctuations in supply and demand that will need to be accounted for on the small scale, resulting in an increase of the overall "cost" of an ETM scenario. These datasets comprising of hourly data spanning one year are also inserted into the graph to be balanced in one calculation. One interesting facet of the ETM is that because it does not linearly progress through time, it is able to adjust values from different points in hourly datasets. In the Energy imbalance mitigation section, the method *forecasting* makes use of this feature. If the model finds an hour of undesirable peak load, it looks back up to 72 hours for any possible excess load that it could store and allocate in the peak load hour. In a way it does not use forecasting, but rather 'backcasting'. This is one feature of the ETM that is possible through not using timesteps.

4.1.3. Space specific information

There is no direct representation of space in the ETM. Data containing attributes of different locations is used as inputs (Quintel, 2022c), but no spatial component exists in the input sliders for the users or the various outputs that the ETM provides.

4.1.4. ETM audit answers summary

Table 4.1: ETM model audit summary

Category	Audit question	Answers (ETM)
general information	What was the initial purpose of the model?	Real-time exploration of future energy scenarios
	What A/D efforts have already been made to the outputs of the model?	Main outputs are aggregated statistics over all sectors. Some hourly data can be exported
	On which modelling principles (and/or locality) is the model based?	Network graph balancing. For each mutation of inputs, the underlying graph is balanced from demand to supply
	Is the model deterministic?	Yes
	What is the percentage of model in/outputs to be mapped vs total in/outputs?	1%
Time specific	How is time represented in the model?	Time is absent. The ETM graph is converted directly from start values to end values.
	What are the timestep, timescale and total runtime?	No timestep. Timescale able to simulate up to 30 years. Runtime <10 sec (by design)
	Through which timeframe is the model expected to communicate?	Only at start and end
	At what frequency is the model expected to communicate?	Once per run
	When is the model expected to communicate?	There are no triggers during a run that could cause communication to occur
Space specific	How is space represented in the model?	No representation of space
	Does the model use geometric map-based data structures?	No
Object specific	How is the level of detail distributed across the model?	Detail is spread relatively evenly. Parameter-wise more detail was given to modelling industry
Sensitivity specific	How sensitive is the model to the inputs that it needs the other model(s) outputs?	not applicable, model will receive no inputs
	How sensitive is the model to variations in input resolution?	not applicable, model will receive no inputs

4.2. HWP model audit

The following subsections highlight facets of the HWP model by category, based on answering the questions in table 3.1. Figure 4.2 shows the HWP model representation in Linny-R. In short, the model and its corresponding language functions as follows:

Linny-R works with sources from which things can be drawn (at a specified price, sinks in which things can be put (at a specified price), storages and processes. Processes (square elements in figure 4.2) can draw and output to and from sources, sinks and storages at will, only constrained by their installed capacity to do so. The windfarm actor attempts to maximize profits by selling electricity to the grid. A set schedule of selling power has been given to the process ”leveren aan stroomnet”, meaning that the model is unable to change how much it will output because of contractual obligations. Power is generated by the process ”windstroomproductie”. This process also has a set schedule of producing power, based on the weather and the assumption that windmills will always generate power when able to do so. There exists a slight weather-induced imbalance between what is generated by ”windstroomproductie” and what ”leveren aan stroomnet” expected to generate. This imbalance will have to be compensated for by the Windfarm actor, either by buying and selling on the expensive imbalance market or by means of buffering with hydrogen.

4.2.1. General information

The HWP model aims to alleviate the imbalance between the agreed quantity of electricity sold on the day-ahead market with the actual production by means of an internal hydrogen buffer. Electricity production is not an output in the HWP, but an input. The only outputs are monetary, due to the stated

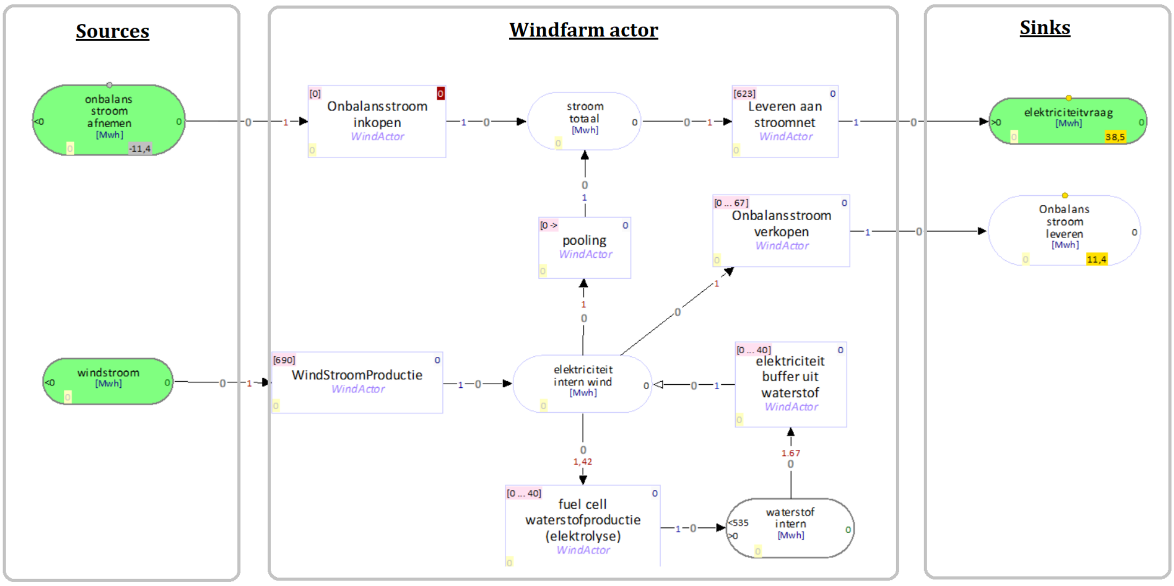

Figure 4.2: Overview of the HWP model in Linny-R

financial objective of the model. Under various specifications of the hydrogen buffer, the model is able to compare marginal cost gains (in €/kWh). The HWP model operates on the base principles of linear optimisation, which revolves around finding an optimum in a solution space bound by decision variables, objectives and constraints (Schulze, 2000). Linear programming models can be seen as a set of linear boundaries in which a space exists that can be maximized to a desired outcome. For example, figure 4.3 shows a linear optimisation problem with two dimensions X and Y. Given the constraints, represented by the lines and the positive axis in this case, a solution space occurs. Depending on the goal function, an optimum can be reached. For example if the X axis represents profit and the Y axis represents CO2-generation, the goal function might be to maximize X and minimize Y at corresponding weights. Depending on the weights of goal X and goal Y, the solving program in figure 4.3 will most likely select point B or C as the optimum, thus completing the simulation.

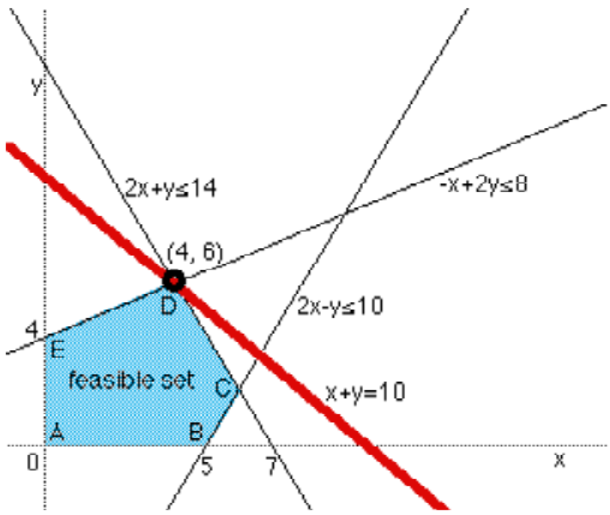

Figure 4.3: linear programming example, from Schulze (2000)

The HWP models' main constraints are the quantities of electricity that it has to produce from wind and the quantities of electricity that it has to supply to the electricity market. Decision variables revolve around balancing the uneven supply and demand through costly imbalance purchases or hydrogen storage. The goal function is singular: maximize profit. Additional goal functions could for example be the maximisation of hydrogen production, but this is not the case.

The model is fully deterministic. No randomness exists in the model, which makes each run with the same input variables result in exactly the same outcome, given a unique optimal solution. This affects coupling, because the HWP model will not have to run multiple repetitions of the same inputs to acquire a stable solution.

Three parameters (settings for the hydrogen buffer) and 3 time series (production, regular price and imbalance price) are used as variable inputs. Out of these, two time series (production and regular price) will be inserted by the ETM. This results in 66% of time series inputs and 33% of total inputs of the HWP model to be coupled. Given that electricity price is the foundation of the only optimisation goal in the HWP, it is reasonable to state that careful consideration has to be given to the A/D functions responsible for the transformation of this important dataset.

4.2.2. Time specific information

Time is handled in this model in discrete intervals of 60 minutes each. The optimisation program receives information regarding prices every new timestep one at a time and optimises profit based on that information. Time in this model is therefore something to integrate over. There is no possibility of any instants (Nance, 1981) occurring between the timesteps. What can however happen is an optimisation over multiple timesteps. The HWP model contains a method of storage: a hydrogen storage tank. This would enable the model to transfer energy deficits or surpluses partially to other timesteps where the opposite problem may be happening. However, since the model initially only maximizes per 1 timestep, it is impossible to plan strategically in advance without additional information. This is why the model also introduces a look-ahead function. The look-ahead function provides the optimizer information regarding the upcoming *12* timesteps, which allows for optimisation over a span of 12 hours at a time instead of 60 minutes.

The model is expected to communicate only before a run starts, not during a run. The HWP model is built on the assumption that the electricity price is not affected by the production of the windfarm itself. Therefore, introducing a feedback during model runs would be going against how the model was initially conceptualised.

4.2.3. Space specific information

A representation of space is not present in the model. However, some sort of spatially distributed data was used (in aggregated form) in order to determine the production of the windfarm itself. The model used daily average wind speeds in multiple measuring locations off the coast in the Netherlands to create a homogeneous wind speed for ”offshore NL”. In this aggregation, the diversity in wind speeds at certain locations is lost in favour of one average. Secondly, the HWP model originally disaggregates the daily data into hourly intervals using equation 4.1.

$$V_{i,j} = \frac{G_j * (1 + random(-\frac{M_j}{G_j}, \frac{M_j}{G_j})) + V_{i-1}}{2} \tag{4.1}$$

$$V_{average,day(j)} = G_j$$

$$V_{max,day(j)} = M_j$$

$$V_{wind}\ on\ t_{houri,dayj} = V_{i,j} = 0 <= V_{i,j} <= M_j$$

4.2.4. Object specific information

Since the HWP model is fairly small in terms of variables and processes, the detail in the model is distributed fairly evenly. Processes have uniformity of possible actions and constraints and are centrally driven by a single mathematically defined goal.

Wind farm power production occurs in both the ETM and HWP. In the HWP model, wind farm power production is models as pre-determined production of electricity (MWh) per timestep (MWh) as an input. The model displays cost and price alternative strategies in the solution space if any exist.

4.2.5. Resolution specific information

The HWP contains 14 variables and 4 time series. 6 of these (3 variables, 3 time series) fluctuate per scenario. The HWP model is set to receive two types of inputs from the ETM:

- Electricity prices (hourly). These will be the input for the parameter 'electricity prices hourly' (Eprice) in the HWP
- Full load hours (yearly). This will be the input for the parameter 'power generation hourly' (Power) in the HWP

Inputs sensitivity

A rudimentary sensitivity analysis was performed by varying the current values of 'electricity prices hourly' and 'power generation hourly' by 10% up, down and neutral. The results displayed in figure 4.4 show the relative change to the output variables 'cashflow' and 'TMK' compared to the base case in the middle. TMK represents the added value per unit of hydrogen produced in the windfarm.

Cashflow	Power +	Power o	Power -
Eprice+	1.21188	1.10303	0.99406
Epriceo	1.09854	1	0.90133
Eprice-	0.9852	0.89697	0.8086

TMK	Power +	Power o	Power -
Eprice +	1.00081	1	0.99925
Eprice o	1.00081	1	0.99925
Eprice -	1.00081	1	0.99925

Figure 4.4: HWP input sensitivity analysis

Figure 4.4 shows that varying the electricity price and power output has a meaningful and significant effect on the revenue generated by the windfarm. The only invariance in cashflow occurs when price and cashflow are changed to the opposite, one +10% and one -10%. Therefore, both inputs seem to affect the revenue on a similar level. However, the added value per unit of hydrogen (TMK) seems invariant to price, but not to power output. This points to the fact that price and power output are very different inputs when it comes to this linear optimisation model. Changing the price does not change the constraints of the HWP. Changing the power output does. As stated in 4.2, the HWP model is forced to produce as much as it can. Changing the power output of the windfarm therefore changes these constraints.

ΔCashflow	Power +	Power o	Power -
Eprice+	1.03482	1	0.95941
Epriceo	1.03482	1	0.95941
Eprice-	1.03482	1	0.95941

Figure 4.5: Windfarm cashflow with and without a hydrogen buffer

The TMK output seems to be reliant on the constraints in the model, but not the electricity price itself. Figure 4.5 confirms this. TMK is calculated in the HWP using equation 4.2.

$$TMK = \frac{Revenue_{with\ hydrogen\ buffer} - Revenue_{without\ hydrogen\ buffer}}{Hydrogen\ produced} \tag{4.2}$$

In the HWP, hydrogen is produced to offset any imbalance in production. If 1MWh fewer is produced than is needed, the hydrogen buffer activates. These imbalances are created by the constraints of production, meaning they stay the same regardless of price. The lower half of equation 4.2 is therefore invariant to price. The upper half of the equation calculates the added value in revenue of using a hydrogen buffer. The delta in revenue of installing such a buffer is also invariant to price, which can be seen in figure 4.5.

Coarseness sensitivity

The inputs from the sensitivity analysis were aggregated and subsequently disaggregated in the time dimension in order to explore how a coarser input would affect model performance and behaviour.

Since both inputs were hourly time series, the same A/D function could be used on both. The A/D that was done uses the widely used equations 2.1 and 2.2 to aggregate the original hourly data to daily averages and then to create hourly values again.

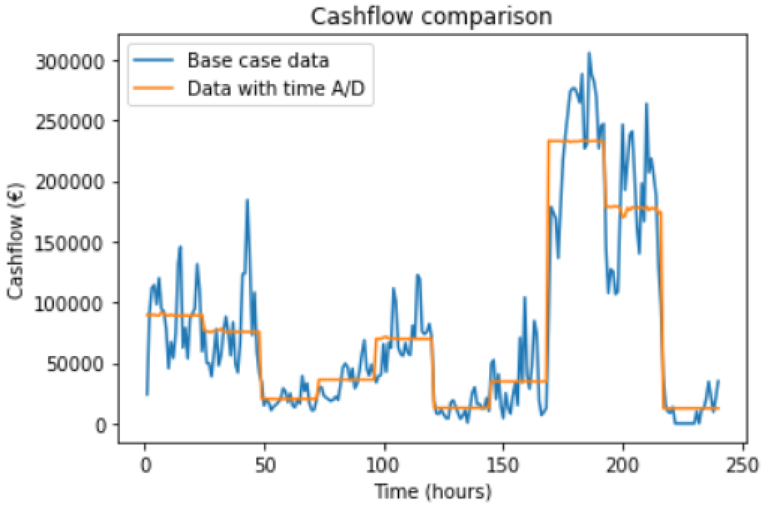

Figure 4.6: HWP cashflow behaviour comparison between original and coarse data

Figures 4.6 and 4.7 compare the behaviour of the model during a run with a regular and a coarsened input. Cashflow progression through time follows a similar pattern over time. Disparities do exist, such as around timestep 25 and 80-100 in figure 4.7 where the hydrogen production contrasts between runs. When the same sensitivity analysis is performed using the coarse datasets, similar results emerge as with the original data. More figures regarding this analysis can be found in appendix A.1.

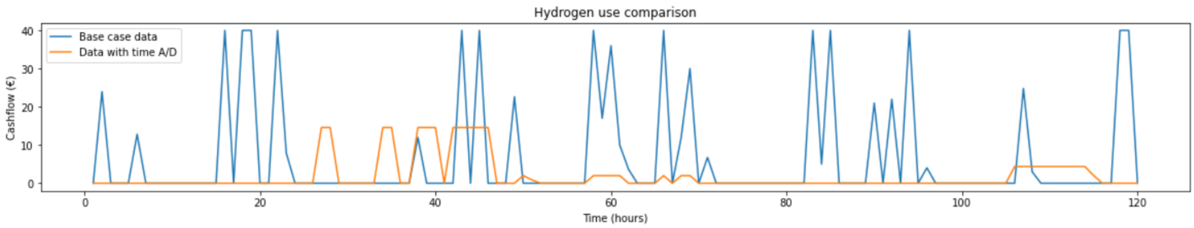

Figure 4.7: HWP hydrogen use behaviour comparison between original and coarse data

4.2.6. Audit answers summary

Table 4.2: HWP model audit summary

Category	Audit question	Answers (HWP)
general information	What was the initial purpose of the model?	Quantify profitability of alleviating the imbalance between wind energy sold and actual daily output with a hydrogen buffer
	What A/D efforts have already been made to the outputs of the model?	Profits are aggregated across time to provide a final profit per year
	On which modelling principles (and/or locality) is the model based?	Linear optimisation. Within the boundaries of what the windfarm has to produce, the model maximises profit with a hydrogen buffer
	Is the model deterministic?	Yes
	What is the percentage of model in/outputs to be mapped vs total in/outputs?	33%
Time specific	How is time represented in the model?	Discrete intervals of 60 minutes. The solver maximises the profit integral from the current step n to n+12 via a look-ahead function
	What are the timestep, timescale and total runtime?	timestep 60 minutes, timescale 1 year, runtime +- 10 minutes
	Through which timeframe is the model expected to communicate?	Only at start and end
	At what frequency is the model expected to communicate?	Once per run
	When is the model expected to communicate?	There are no triggers during a run that could cause communication to occur
Space specific	How is space represented in the model?	No representation of space.
	Does the model use geometric map-based data structures?	Spatial input data from offshore wind-measuring stations was aggregated in the time and space dimension to one value per hour
Object specific	How is the level of detail distributed across the model?	Detail is distributed evenly
Sensitivity specific	How sensitive is the model to the inputs that it needs the other model(s) outputs?	Revenue is (equally) sensitive to both expected inputs. Model output 'TMK' only sensitive to the constraint-based input 'power production'
	How sensitive is the model to variations in input resolution?	Output KPI's are not significantly affected by coarsening of input data. Behaviour on individidual timesteps is altered

4.3. EVM model audit

The following subsections highlight facets of the EVM model by category, based on answering the questions in table 3.1.

4.3.1. General information

The EVM is an agent-based model used to explore how electric vehicle (EV) power demand *and* EV battery available capacity are distributed in time and space in the Netherlands. The model is based on the modelling principles of agent-based modelling and discrete time step simulation (DTSS). A population of agents is created that interacts with it's surroundings based the agents' internal state and the state of the surrounding environment. Each timetick, the agents perform their actions one by one in a seed-based random order. The environment in this ABM concerns mainly the price of electricity at a given timestep. All EV-agents go through their daily travel routines travelling from A to B. When stationary, the EV-agent is able to charge. How much they charge and where this charge is drawn from depends on the location and the current prices. Aggregations to the output have been already made in this model. The electricity demand and battery availability for all vehicles is aggregated spatially to municipality level each timestep.

Randomness is a part of the EVM both during setup and during runs. Random distributions are used to set the starting attributes for EV-agents such as home and work location, battery size, whether the agents charges 'smart' or not. During the run, random distributions are used to offset scheduled departure and arrival times. The case 2 coupling will couple two inputs of the EVM. This translates to 7% of the 29 total inputs that will be reliant on the coupling. While this is definitely less than the HWP, one of these inputs that is received from the ETM (price) is the EVM's only time series and not to mention a vital driving factor behind the behaviour of the EVM.

4.3.2. Time specific information

DTSS models use a time tick with a constant interval between ticks to progress time. In the case of the EVM, each tick represents 15 minutes. Between these ticks, nothing is able to occur. Some actions that EV-agents can do are not done in 15-minute intervals, such as the time it takes to drive from point A to point B. A more disaggregated time tick could have increased the resolution of this action further, but this was not done. The model contains no look ahead, but EV-agents do have a memory of past electricity prices. Based on this memory, the agents attempt to predict when prices will be lowest in the future in order to charge at the cheapest times. When prices are predicted to be the same, an EV agent will choose the earliest next moment to charge. The model is run for 672 timesteps (1 week in system time). Communication is assumed to occur only at the start and the end of a run.

4.3.3. Space specific information

EV agents are assigned a location of residence and a location of work at the start of the model. The aggregated input that determines the total population of agents is spatially disaggregated to achieve this. The EVM divided the Netherlands into municipalities which have been aggregated to a centroid per municipality, as seen in figure 4.8.

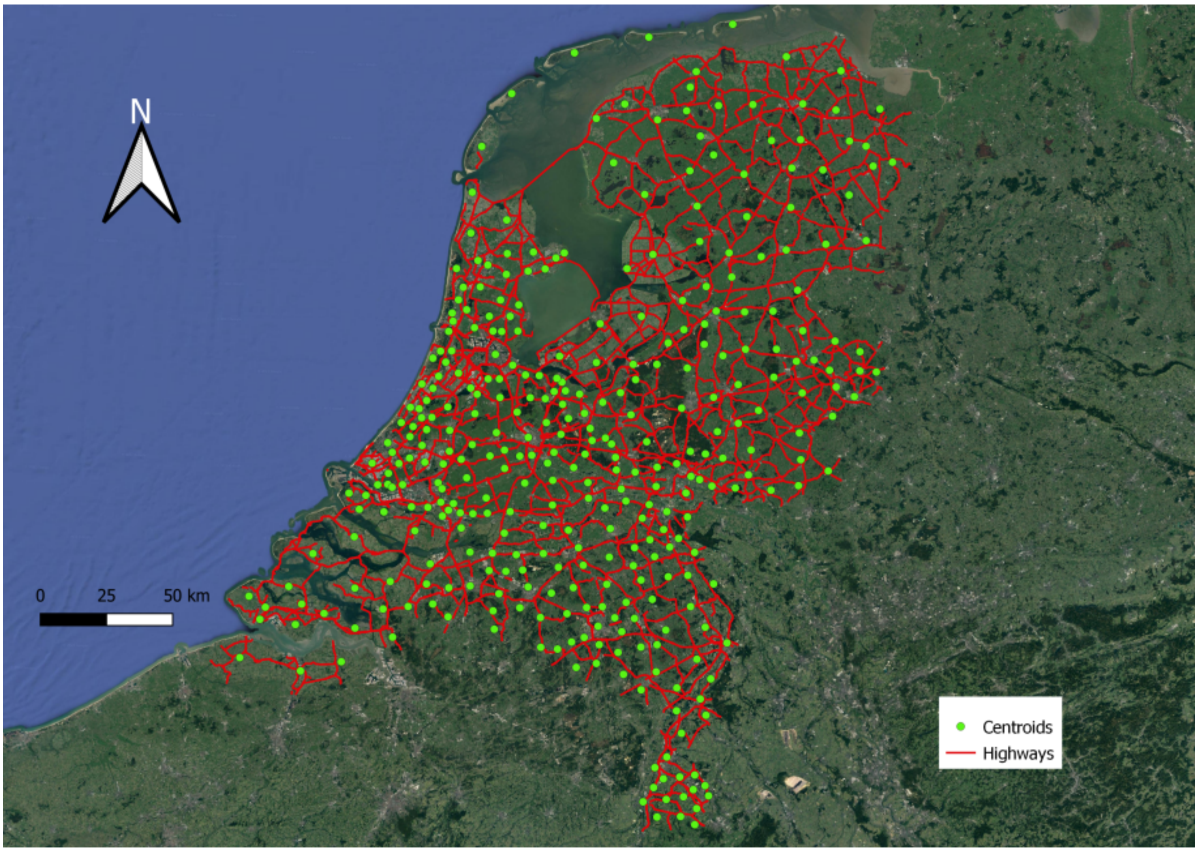

Figure 4.8: EVM roads and municipality centroids

The centroids were based on the center of mass of each municipality, using existing polygon data. Each EV agent is therefore only able to travel from centroid to centroid, reducing the spatial resolution of the model. Driving distance to other municipalities (other centroids) was implemented using QNEAT

3, which is able to approximate distance between two points via road networks. Initial agent population is also attributed per municipality at the start of the model

4.3.4. Object specific information

The EV-agent has the focus in the EVM. Most of the detailed behaviour is attributed to this agent. However, since the EV-agent is the only agent in the model, this allocation of detail could be expected. However, within the agent the level of detail also varies. Much attention is given to the mechanisms by which the agent determines when to charge. Limited attention is given to how the agent decides to travel every day. For example, the EVM assumes a five-day workweek every week of the year, no vacation trips and no non-work related (weekend) travel. This disparity in detail creates quite homogeneous schedules for all agents. They might still all use complex methods for determining when to charge, but where and how much is determined by limited travel behaviour.

4.3.5. Resolution specific information

The EVM has a total of 30 input variables, totalling 28 float inputs, 1 timeseries for the electricity price and 1 dataset containing population counts per municipality. The EVM model will receive two types of input from the ETM:

- Electricity prices (hourly), which will be the input for the variable 'electricity_price' (quarter-hourly) of the EVM.
- Population (national), which will be the input for the variable 'EVs_per_municipality (municipality level) of the EVM

Inputs sensitivity

The same rudimentary sensitivity analysis was performed as for the HWP in section 4.2.5. The respective inputs of total population size and electricity price were varied 10% up, down and neutral. The results are displayed in figure 4.9.

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.899
Epriceo	1.099	1.000	0.899
Eprice-	1.099	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.900
Epriceo	1.094	1.000	0.900
Eprice-	1.094	1.000	0.900

Figure 4.9: EVM input sensitivity analysis

The KPI's in figure 4.9 seem invariant to price. Something which could be explained by an observation done in section 4.3.2. Section 4.3.2 states that EVs select the cheapest prices to charge. That statement combined with figure 4.3.2 implies that EVs are invariant to sensitivity of price, because it's the relative cheapness of prices that matter. The agents will have to charge at some point. If e.g. step 6 was the cheapest and all prices are increased by 10%, step 6 would still be the cheapest. Therefore behaviour is unaffected by increasing price by a fixed percentage. Varying the amount of EVs almost perfectly translates to changes in the KPI's. A variation of 10% up or down results in a respective KPI movement of roughly 10% as well.

Coarseness sensitivity

While increasing the price by a fixed percentage does not affect behaviour, changing the price data itself does. Figure 4.10 shows the comparison of a normal run to a run where price data has been coarsened by setting the price of each 15-minute interval to its daily mean value. By removing the daily fluctuations that the EV-agents react to, the behaviour shifts from strategic charging to charging whenever they need to. A clear indicator of this shift is the disappearance of the various peaks around t=20, t=115 and t=210. At these times, the price of electricity will be at a local minimum, thus attracting any stationary cars to charge if needed. Now charging occurs more gradually, whenever cars arrive at a charging station. The 'relative cheapness' of price data will therefore be important to consider when creating A/D functions.

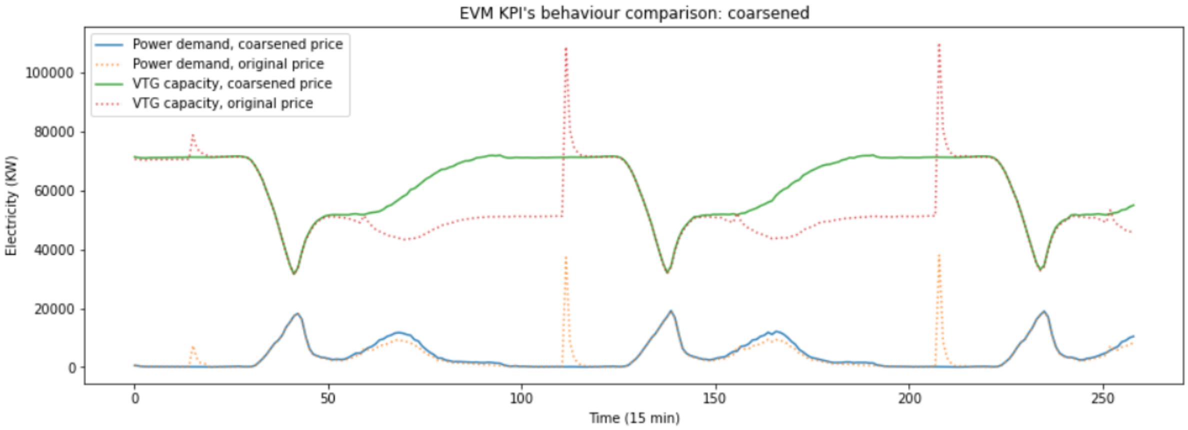

Figure 4.10: EVM behaviour compared to using coarsened data

The model behaviour, while altered by coarsening, responds similarly to the sensitivity analysis done on the original price data (see figure 4.11).

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.899
Epriceo	1.099	1.000	0.899
Eprice-	1.099	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.900
Epriceo	1.094	1.000	0.900
Eprice-	1.094	1.000	0.900

Figure 4.11: EVM input sensitivity analysis, coarsened price data

4.3.6. Audit answers summary

Table 4.3: EVM model audit summary

Category	Audit question	Answers (EVM)
general information	What was the initial purpose of the model?	Explore how electric vehicle (EV) power demand and EV battery available capacity are distributed in time and space in the Netherlands
	What A/D efforts have already been made to the outputs of the model?	Output power demand and battery availability per EV agent are aggregated spatially to a municipality level
	On which modelling principles (and/or locality) is the model based?	The model is agent-based and uses DTSS conventions such as a timetick and a schedule of agents to perform actions
	Is the model deterministic?	No. Randomness is used before model run to set up initial agent attributes and during model run to alter certain values
	What is the percentage of model in/outputs to be mapped vs total in/outputs?	7%
Time specific	How is time represented in the model?	Discrete intervals of 15 minutes. Agents act one by one each timestep. Agents have a memory to store past system states (electricity prices) to attempt to predict future prices with
	What are the timestep, timescale and total runtime?	timestep 60 minutes, timescale 1 year, runtime +- 10 minutes
	Through which timeframe is the model expected to communicate?	Only at start and end
	At what frequency is the model expected to communicate?	Once per run
	When is the model expected to communicate?	There are no triggers during a run that could cause communication to occur
Space specific	How is space represented in the model?	Municipalities are represented as centroids of their corresponding polygons. Agents travel between centroids via existing road networks in the Netherlands. From a single starting agent count, agents are spatially distributed to live in different municipalities.
	Does the model use geometric map-based data structures?	Polygons for municipalities are used, alongside spatial data regarding highways in the Netherlands
Object specific	How is the level of detail distributed across the model?	Detail is concentrated in the price choice functions of the EV agent. EV travel behaviour, which drives charging demand, is modelled much less detailed
Sensitivity specific	How sensitive is the model to the inputs that it needs the other model(s) outputs?	Both KPI's are sensitive to a larger EV-agent population and insensitive to altering price by a fixed percentage
	How sensitive is the model to variations in input resolution?	Lower price-seeking behaviour is dulled when price resolution is reduced

4.4. Case 1 coupling audit

4.4.1. Coupling pairs overview

As seen in figure 3.1, the ETM will provide two inputs to the HWP. An overview of the input and output units can be found in table 4.4. Immediately, pair 2 is notable because of its apparent fit. The timeseries outputted by the ETM has the correct shape and unit to be directly inserted into the HWP. However, even though the shape and units are already aligned, the semantics will also have to cohere to a sufficient degree. Pair 1 is less well fitting from the start, with units and shapes both being misaligned. While the ETM is significantly larger in terms of inputs compared to the HWP, it seems that the data from the ETM is at a lower resolution level than the smaller HWP. This contrasts with Davis and Bigelow (1998a), which states that the model with the larger parameter count usually has the higher detail level. However, this statement does not factor in that the ETM has a much larger scope than the HWP. When it comes strictly to the parameters that are of interest to the coupling, the ETM has a lower resolution.

Table 4.4: ETM-HWP variables unit comparison

Input pair number	Parameter name	Model of origin	Input or Output	Parameter unit	Parameter shape
1	Yearly full load hours, Wind speeds	ETM	Output	Hours/year, Decimeters/second/day	- 1 float number (h/y) - 1 timeseries, 365 entries (dm/s)
	Wind speeds for power generation	HWP	Input	meters/second	- 1 timeseries, 8760 entries (m/s)
2	Electricity price	ETM	Output	€/MWh	- 1 timeseries, 8760 entries (MWh/h)
	Elektriciteitprijs	HWP	Input	€/MWh	- 1 timeseries, 8670 entries (MWh/h)

The first output-to-input pair does not seem as well-aligned. This is true, because the HWP and ETM both use wind patterns to calculate power output after the data has been inserted into the model. The HWP model converts the inputs of wind (U) and full-load-hours (F) it receives into MW power (P) using equations 4.3 and 4.4. However, both the units and the time resolution are different between the input and output.

$$P = \begin{cases} 0\ if\ U \leq 3\ m/s \\ 0\ if\ U \geq 25\ m/s \\ P = cU^3\ otherwise \end{cases} \tag{4.3}$$

$$P \leq P_{setup}$$

$$c = \frac{P_{setup} * F_{full\ load\ hours}}{\int_{t=0}^{T} P_{t,\ total}} \tag{4.4}$$

4.4.2. Semantics check

Due to the disparity in time resolution and units used, a disaggregation effort will need to be done in section 4.4.3. But first the semantics of the inputs and outputs will need to be checked. Both the ETM and HWP use space-aggregated wind data from the KNMI to use as input for their models. This means that wind is conceptualised in both models as a directionless vector of speed. Full load hours of some sort are used as a calibration tool in both models in order to scale the power production calculated like in equation 4.3. The HWP uses CBS data to calibrate equation 4.3, but the principle overlaps enough to be semantically transferable.

Price in the HWP model is semantically defined as the price that the wind actor can receive for the sale of a MWh of energy to the market, represented by the sink 'electriciteitvraag' in figure 4.2. In the ETM however, price is constructed via a merit order of multiple different types of power generation. Each type of energy producer (coal, wind, gas, etc.) is conceptualised to have a set price at which they will start producing. Wind power also has such a price in the ETM. The HWP does not use a merit order concept, but strives to produce whenever possible to minimize loss of green energy produced. In fact, the constraints in the HWP model deny any option of not producing, regardless of price. If the HWP viewpoint has to cohere with the ETM, one can say that the HWP has a merit order production starting price of $-€\infty$. This coheres with the way the HWP behaves and with the ETMs view on pricing.

As for pricing data itself, it could possibly produce values that would nullify the use of the HWP. Regular negative prices due to possible oversupply of energy in the future may break the model, as a negative cashflow from the start makes for no viable business strategy. The model will still work, but the result will be a resounding 'no' for the company. Aside from that, while prices are being given for a new year as an input, the prices for imbalance remain the same. The HWP model assumes that the cost of repairing imbalance is structurally more expensive than the revenue generated from producing electricity. If this relation were to flip due to significant price changes in the ETM, the HWP model could be placed in an unrealistic state where it would be more profitable to pull the network out of balance than to have it remain in balance. This relation should be monitored carefully when making the coupling.

4.4.3. A/D function requirements

The disaggregation of the ETM wind speed data will transform the data from a timeseries of 365 datapoints to a timeseries of 8760 datapoints. These sizes are fairly manageable by standard python functions. There is little expectation that the size of the dataset alone will result in problems during disaggregation.

System specific constraints

$$E_{thresh} * \frac{\sum_{i=1}^{N} Wind\ speed_{disagg_i}}{N} = Wind\ speed_{agg} \tag{4.5}$$

An equal average rule can be implemented like in equation 4.5. The A/D is only expected to occur once, before the run starts. The risk of thrashing is likely low under these circumstances. Therefore a margin of error E_{thresh} would be acceptable to use. The error threshold can be viewed as the induced translation error during disaggregation. Of course, a lower margin of error between the aggregated and disaggregated states remains a goal to strive for.

Logic specific constraints

When looking at windspeeds, there is a solid foundation of research to aid in disaggregation efforts. Wind speeds generally follow a weibull distribution (DWIA, 2003). Figure 4.12 shows that the mean wind speed data received from the ETM follows such a weibull distribution as well. A disaggregated dataset, by whichever means it is done, is therefore expected to follow a weibull distribution too. Adhering to the distribution can be a vital pointer to the creation and testing of proper disaggregation functions in this case.

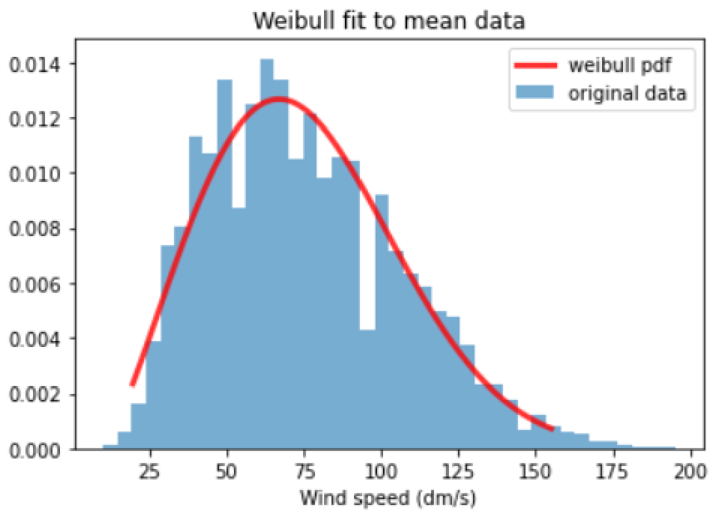

Figure 4.12: ETM HWP Weibull fit to mean data

Furthermore, the dataset from the ETM comes with two additional datapoints: *the maximal and minimal wind speed measured for every day*. These set the domain limits of values that have been recorded on those days and can therefore also function as boundaries for the disaggregation function. Note that these boundaries vary day to day and that they do not affect the general pattern displayed in figure 4.12.

4.4.4. Consistency maintenance

Consistency checks for this disaggregation are unable to compare the behaviour of the model using the aggregated and disaggregated dataset, because the HWP is only able to complete a run with the disaggregated data. Therefore consistency cannot be checked by comparison of error between using both datasets. However, drawing from the viewpoints mentioned in 2.5.1 and 2.5.2, consistency in the wind speed disaggregation can be interpreted as adhering to the following two principles:

1. The disaggregation effort must adhere to the logic specific and system specific constraints of the aggregated dataset
2. Instances of disaggregated data must produce similar behaviour within a margin of error

The 'correct' margin of error is difficult to determine and ultimately comes down to the tolerance of the model's users. Margins of error also occur in multiple places, such as in comparing output behaviour but also in equation 4.5. Although these margins are arguably correlated due to the tightness in error in equation 4.5 directly influencing the variety of inputs to the model, they are not exactly the same. For the purpose of this hypothetical case, it is assumed that both the model's users aim for margins of error to be at a feasible minimum at both cases. The feasible limit would then lie where any reduction in error margin would result in such an increase in static or dynamic transformation and consistency cost (2.5.3), that it would not be viable to do so.

In addition, A/D functions with lower margins of error are more able to withstand thrashing-related drift. Any A/D effort with a margin of error can eventually drift away given enough cycles of A/D. However, the drift can be lowered by reducing the initial allowed margin of error or by reducing the frequency of subsequent A/D. The case 1 coupling fortunately will not suffer much from thrashing, because the A/D effort only occurs once per run. Nevertheless drift away from the original data introduces error which, as discussed in the previous paragraph, should be kept low whenever possible.

4.4.5. Case 1 coupling audit summary

Table 4.5: ETM-HWP variable specific audit summary

Category	Audit question	Answers (ETM-HWP)
Coupling pairs overview	What are the inputs and respective units required for the coupling?	See table 4.4
	For each input, what is the corresponding output and respective unit in the other model?	See table 4.4
	Are the semantics, syntactics and pragmatics of each input/output pair consistent?	Yes, semantics for price and wind data have sufficient overlap to be coupled
	What levels of resolution can the models be described as having?	The wind output of the ETM will require disaggregation. The HWP has the higher resolution in this coupling, due in part to the broadness of the ETM.
A/D function requirements	What are the system specific A/D constraints present for each input/output pair?	Price requires no change in resolution. Wind speed is uneven in the temporal dimension and could be handled using an equal average rule with margin of error
	What are the logic specific A/D constraints present (if any) for each input/output pair?	Wind speed data adheres to a weibull distribution in literature and in the ETM-supplied data. Further data regarding min and max can be used as constraints for disaggregation
	Are there input/output pairs with constraints that are sufficiently similar that an A/D function could be re-used?	No. Only one disaggregation is present. Therefore no re-use within this coupling is possible
	What are the estimated input/output data sizes?	See table 4.4
	What is a realistic domain for each input and output?	Maxima and minima have been provided. A natural constraint on wind speeds is also that no wind speed can be <0 m/s
Consistency maintenance	What is an adequate goal of consistency to strive for in this model coupling?	- The disaggregation effort must adhere to the logic specific and specific constraints of the aggregated dataset - Instances of disaggregated data must produce similar behaviour within a margin of error
	What is an acceptable margin of error between aggregated and disaggregated values?	it is assumed that both the model's users aim for margins of error to be at a feasible minimum at both cases
	Are the A/D functions sufficiently able to withstand thrashing?	A/D occurs infrequently enough that thrashing is not a direct threat to consistency
	What are the static and dynamic consistency costs?	TBD during experimentation

4.5. Case 2 coupling audit

4.5.1. Coupling pairs overview

Figure 3.2 shows which two outputs the ETM provides to the EVM. The accompanying inputs and units of the EVM are displayed in table 4.6. Input pair 1 is misaligned in terms of spatial resolution. The input of the ETM only supplies total population in the Netherlands during a given year, while the EVM requires a specific amount per municipality. In addition, the ETM supplies a count of people in general, which needs to be transformed to EV-agents (persons with an EV). Input pair 2 is misaligned in the time dimension. A time disaggregation will be needed to transform the ETM's hourly values in pair 2 to quarter-hour values. Similarly to case study 1, the ETM is significantly larger than the EVM, while still having outputs at a lower resolution. The broadness of the ETM compared to the EVM once again causes this disparity.

Table 4.6: ETM-EVM variables unit comparison

Input pair number	Parameter name	Model of origin	Input or Output	Parameter unit	Parameter shape
1	Population size	ETM	Output	Persons	- 1 integer number (persons) - 1 float number (percentage total car use) -1 float number (percentage EVs of total cars)
	EV population	HWP	Input	EVs/municipality	- 352 space-distributed integers (EVs/municipality)
2	Electricity price	ETM	Output	€/MWh	- 1 timeseries, 8760 entries (MWh/h)
	Electricity price	EVM	Input	€/kWh	- 1 timeseries, 35040 entries (KWh/15min)

4.5.2. semantics check

The ETM and the EVM's definition of 'population' varies. Whereas the ETM understands population as being the total population in the Netherlands in any given year, the EVM understands population as the amount of 'EV agents' that it has. EV agents are a combination of an Electric vehicle and its owner into one agent. Fortunately the ETM is also able to provide statistics on the share of cars in total transportation kilometers and the share of electric vehicles of total cars. With this added data, both 'populations' are able to be transferable to each other. It is important to note that this semantic coupling relies on the assumption that e.g. '70% of kilometers are travelled by car' is translated to '70% of the population drives a car (exclusively)'.

The ETM's view on the semantics of electricity price have been previously discussed in section 4.4.2. The EVM takes a different semantic point of view, namely that electricity price is the invariant price that an EV agent has to pay per kWh used at any given 15-minute interval if it decides to charge at that interval. The semantic difference between the ETM and the EVM is expected and logical. The ETM takes a definition of electricity prizes from the perspective of the suppliers. The EVM in turn takes a demand-side perspective. These are not exclusive of each other, but rather two sides of the same coin. The semantics of the ETM and EVM on price can therefore be considered sufficient for coupling.

4.5.3. A/D function requirements

The case 2 space disaggregation requires the allocation of one integer count of population over 352 municipalities. The time disaggregation requires the transformation of a timeseries of size 8760 to a timeseries of size 34050. Given the needed sizes it is expected that data size will not result in problems during disaggregation.

System specific constraints

Interpretations of the equal average rule of equation 2.6 can be found for the spatial disaggregation in equation 4.6, 4.7 and for the time disaggregation in equation 4.8. Do note that the the error thresholds for the spatial and temporal disaggregations need not be the same value.

$$E_{thresh} * \sum_{i=1}^{N} EVagents_{municipality\ i} = EV\ Population_{ETM} * Share\ cars_{ETM} * Share\ EVs_{ETM} \quad (4.6)$$

$$EV\ Population_{ETM} = Population_{ETM} * Share\ cars_{ETM} * Share\ EVs_{ETM} \tag{4.7}$$

$$E_{thresh} * \frac{\sum_{i=1}^{N} Price_{disagg_i}}{N} = Price_{agg} \tag{4.8}$$

Logic specific constraints

Spatially there is ample data available on the distribution of population in the Netherlands. Maps like figure 4.13 already give a clear indication that population distribution is not uniform. A normalisation of these population distributions can be used to disperse the single population number from the ETM across municipalities in the Netherlands. In addition it is logical to state that no municipality can have an EV-agent population below zero or above the total population size.

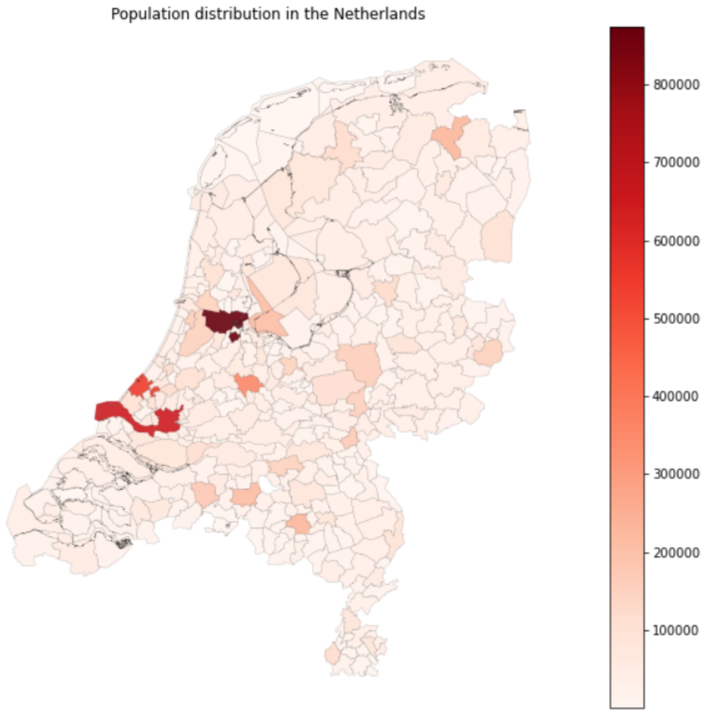

Figure 4.13: Population distribution in the Netherlands (CBS, 2022)

From section 4.3.2 of the EVM model can be drawn that EVs are sensitive to the differences in price over time. The absolute price level does not influence EV behaviour at all, only the differences. Therefore the consistency threshold for price disaggregation was chosen to measure the degree to which a disaggregate captures the change in price, rather than its adherence to the absolute price levels. A full technical breakdown of the reasoning and implementation behind this consistency function can be found in appendix B.1

For electricity prices it is also assumed that energy prices are not allowed to reach below zero. Although recent events show that a zero-threshold for prices can be broken (Bloomberg, 2022), the ETM still upholds this constraint (Quintel, 2021). Given that the EVM is indifferent to such a constraint, the zero-threshold is advised in order to not unnecessarily create conceptual mismatches.

4.5.4. Consistency maintenance

Similarly to 4.4.4, there are a few principles that create a base for consistency maintenance in the case 2 coupling:

1. The space disaggregation effort must adhere to the logic specific and system specific constraints of the aggregated dataset

2. The time disaggregation effort must adhere to the system specific constraints of the aggregated dataset. In particular the price difference from hour to hour (and not the price itself) is important to translate during disaggregation

3. Instances of disaggregated data must produce similar behaviour within a margin of error

In this coupling it is difficult as well to determine a correct margin of error beforehand. Through experimentation and analysis in section 5.4 it will become more clear what the effects of tighter or looser consistency thresholds has on the disaggregation and the coupling itself. If the effects of a tighter threshold are favoured by the stakeholders, the threshold can be tightened. In any case it is important that the effects of consistency maintenance decisions are clearly and concisely documented.

4.5.5. Case 2 coupling audit summary

Table 4.7: ETM-EVM variable specific audit summary

Category	Audit question	Answers (ETM-EVM)
Coupling pairs overview	What are the inputs and respective units required for the coupling?	See table 4.6
	For each input, what is the corresponding output and respective unit in the other model?	See table 4.6
	Are the semantics, syntactics and pragmatics of each input/output pair consistent?	Yes, semantics for population and price have sufficient overlap to be coupled
	What levels of resolution can the models be described as having?	Both outputs of the ETM will require disaggregation. The EVM has the higher resolution in this coupling, due in part to the broadness of the ETM.
A/D function requirements	What are the system specific A/D constraints present for each input/output pair?	- Population requires a spatial disaggregation from national to municipality level. - Price requires a temporal disaggregation from hours to 15-minute intervals.
	What are the logic specific A/D constraints present (if any) for each input/output pair?	- There is ample data on population distribution to use as a base to divide EV-agents across municipalities. EV-agent count per municipality must be above zero and below (or equal) the total population. An equal average rule can be applied to this disaggregation. - Electricity price normally follows a normal distribution, however this is not the case here. Given that EV-agents are only sensitive to *differences* in price, an equal average rule is not the best fit. Instead, an equal difference rule can be used to more effect.
	Are there input/output pairs with constraints that are sufficiently similar that an A/D function could be re-used?	No. There are no two disaggregations in the same dimension (time, space, etc.). The disaggregations present differ to such a degree that difference disaggregation functions and consistency functions are needed.
	What are the estimated input/output data sizes?	See table 4.6
	What is a realistic domain for each input and output?	- 0 ≤ population in a municipality ≤ total pop - electricity price ≥ 0
Consistency maintenance	What is an adequate goal of consistency to strive for in this model coupling?	- The space disaggregation effort must adhere to the logic specific and specific constraints of the aggregated dataset - The time disaggregation effort must adhere to the system specific constraints of the aggregated dataset. In particular the price difference from hour to hour (and not the price itself) is important to translate during disaggregation. - Instances of disaggregated data must produce similar behaviour within a margin of error
	What is an acceptable margin of error between aggregated and disaggregated values?	It is assumed that both the model's users aim for margins of error to be at a feasible minimum at both cases
	Are the A/D functions sufficiently able to withstand thrashing?	A/D occurs infrequently enough that thrashing is not a direct threat to consistency
	What are the static and dynamic consistency costs?	TBD during experimentation

5

Model coupling

Now that both model pairings have been thoroughly researched on a conceptual and a variable-based level, the couplings can be created. This chapter includes the creation of the concrete functions that were used to complete the case 1 and case 2 couplings. The functions are created using the insights gained from chapter 4.

5.1. Case 1 coupling

The main challenge in case 1 is the wind speed temporal disaggregation. In addition, equation 4.3 of the variable based audit exposes that the model calculated windfarm power output by raising the windspeed input to the power of three. Any error in disaggregation will therefore be amplified as well. All the more reason to have tight consistency enforcement methods.

5.1.1. Wind speed temporal disaggregation

With the information present regarding the wind speeds, there are four schools of thought for the disaggregation functions:

1. **A baseline of disaggregation**
   Section 4.2.5 shows a coarse means of disaggregation, where the mean of a day is simply set as the value for each hour in that day. This disaggregation method can function as a baseline to compare other disaggregation efforts.

2. **Weibull-based disaggregation**
   Section 4.4.3 shows that wind speeds, in nature and in the ETM, follow a weibull distribution. A weibull distribution fitted to the original dataset could be used to draw hourly values from.

3. **Triangular-based disaggregation**
   Strict consistency to the provided data is key to prevent the disaggregation functions from being able to influence the multi-model behaviour. With that said, the daily max, min and mean values can also be used to provide triangular-based distributions per day to draw hourly values from. Whether the resulting dataset adheres to the weibull distribution seen in nature and the aggregated dataset remains to be seen

4. **Rolling average (& uniform) based disaggregation**
   The HWP originally also had the problem of having to disaggregate daily wind speeds to hourly values.Originally, a combination of a uniform distribution and a rolling average-based disaggregation method was used. This method, as seen in equation 4.1, was preferred over any single-distribution-based aggregations, because of the smoother wind patterns it provides over time. Distribution-based values have the potential to jump from almost no wind to gale-force wind randomly, as long as for example the average remains consistent with the aggregated dataset. This was deemed unrealistic in the original modelling effort.

Figure 5.1 shows the resulting hourly wind speeds of all mentioned disaggregation techniques except the previously highlighted baseline. In addition, the pdf of the original output data from the ETM is plotted.

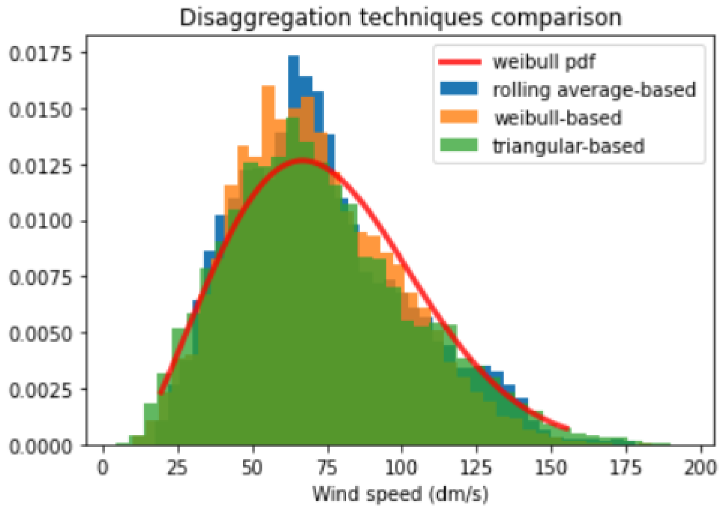

Figure 5.1: Case 1 disaggregation techniques comparison

Figure 5.2 below and figures A.3, and A.5 (in the appendix) show the individual results of the disaggregation functions used, alongside the weibull distribution fitted to the original data. For each disaggregation technique, a kolmogorov-smirnov test was done to test whether the disaggregate data can be assumed to be drawn from the original distribution. Unfortunately, all three techniques failed to confirm the null hypothesis and are thus statistically different from the original distribution. While the triangular-based disaggregation technique had a fit that was several orders of magnitude better than the other disaggregates, it also failed to pass. However, when comparing the figures, the triangular-based disaggregate seems to be the best option so far.

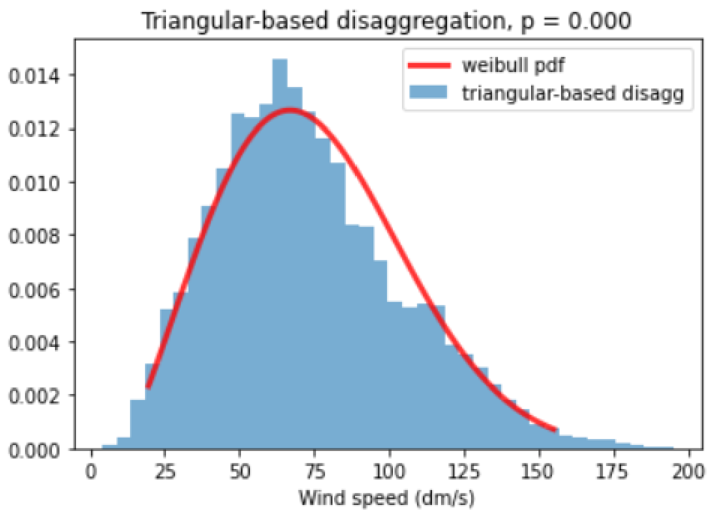

Figure 5.2: Triangular-based disaggregation with accompanying kolmogorov-smirnov test (p-value)

For all disaggregation methods, the wind speeds were converted to the appropriate HWP input (MWh/h) using equations 4.3 and 4.4. The second output from the ETM, the scalar 'Full load hours' is used here as well to create the resulting power curves seen in figure 5.3. The HWP models a setup capacity of 5633MW. Hence the resulting power input curves vary between 0 MW and 5633 MW over time.

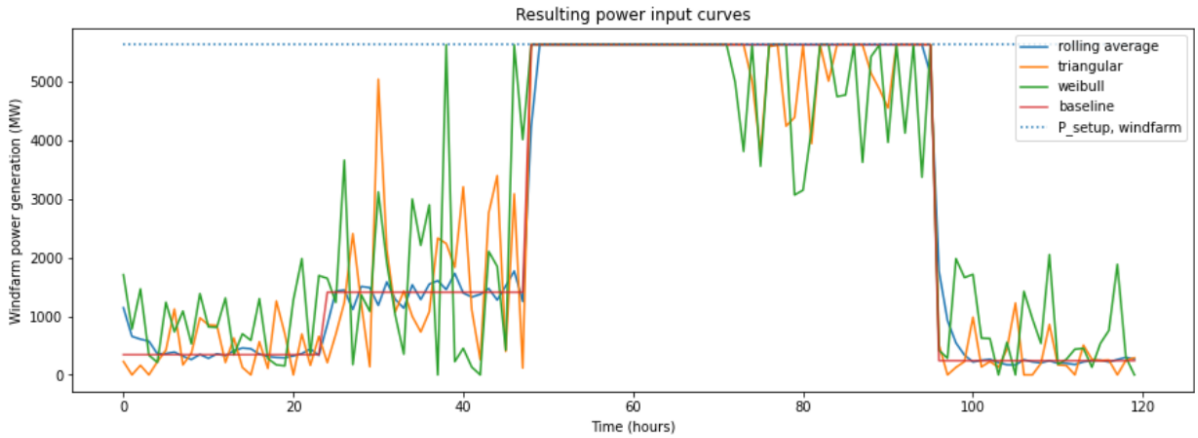

Figure 5.3: Power input curves generated from each disaggregation technique

In figure 5.3 the baseline disaggregation in red can be viewed as the original value of the ETM data, converted to power. The three remaining disaggregation techniques do all revolve around this baseline as expected. The manner in which they do varies quite largely from technique to technique. For example, figure 5.3 shows significant short-term deviations from the baseline with the weibull and triangular disaggregates, while the rolling average technique does not show this at all. This difference is a result of the different ways that the techniques generate wind data, which will be highlighted during verification in section 5.1.2.

5.1.2. Disaggregation techniques verification

The disaggregation techniques presented are meant to generate power input curves varying between 0MW and 5633MW per hour for 8670 hours. A subset of these values is presented in figure 5.3, confirming that this step functions as was intended. The foundation of these power curves is hourly wind data generated from the daily mean, max and min values given by the ETM. Appendix A.3 shows a sample of the disaggregated wind data generated by each technique. The results of the weibull- and rolling average-based disaggregation are also presented below in figures 5.4 and 5.5 for further analysis.

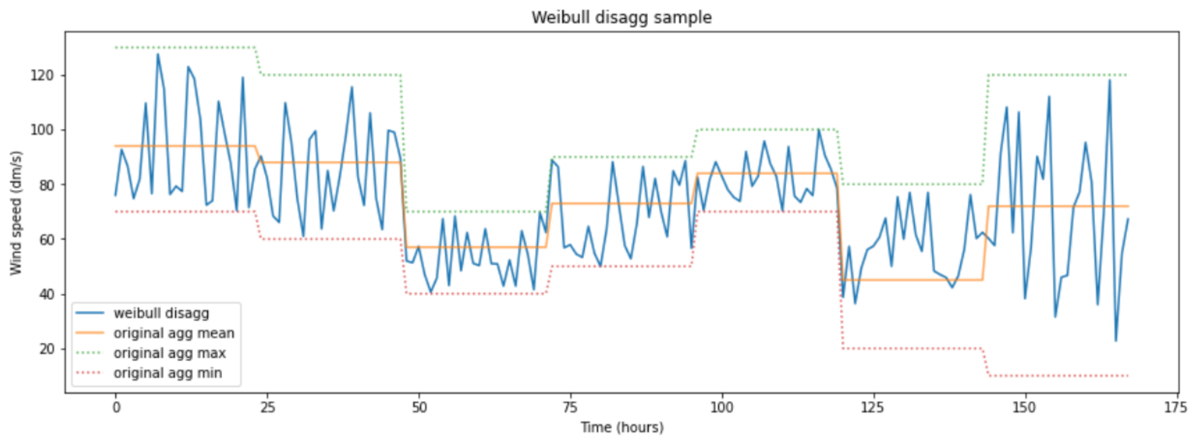

Figure 5.4: One week data sample of weibull disaggregation

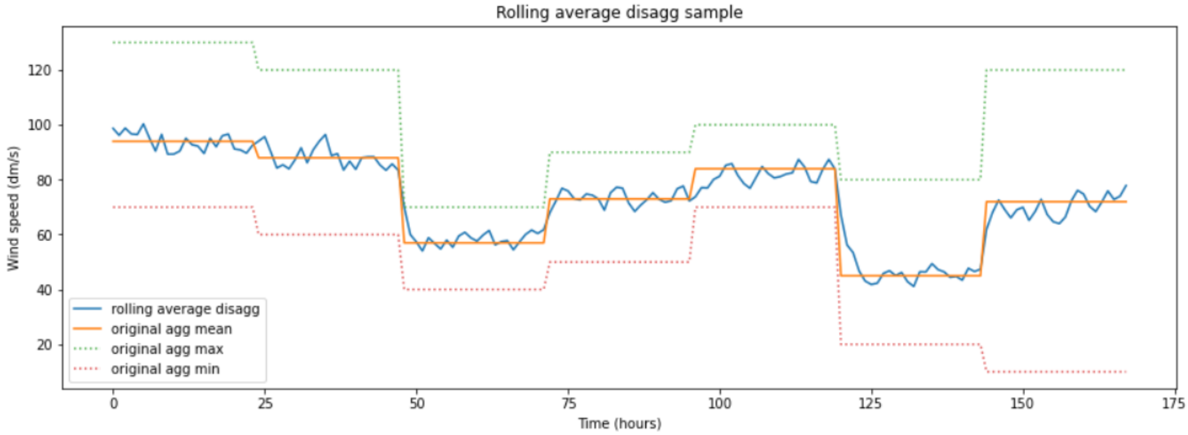

Figure 5.5: One week data sample of rolling average disaggregation

Figure 5.4 and 5.5 show that techniques with relatively similar pdf's can have very different resulting hourly values. The rolling average technique creates wind speeds that stay close to the average. While this can be a good thing, the technique fails to capture the variety in the original data, exemplified by the daily max and min values. Therefore a rolling average technique presents disaggregated wind as more stable as it actually is; something that could affect models very sensitive to fluctuation. On the other hand, the weibull disaggregation captures a much broader spectrum of wind speeds. This also is the case for the triangular disaggregation sample. As further seen in appendix A.3, the samples show that the weibull and triangular disaggregation techniques are able to capture the broadness of wind speeds, while the rolling average and baseline techniques fall short in this aspect.

5.1.3. Coupling run results

The resulting KPI's from running the coupled models with the varied types of disaggregation can be seen in table 5.1. The differing types of disaggregation do not seem to generate meaningfully different run results. The main KPI cashflow only varies 1.6% between the highest and lowest results, with the KPI TMK varying 1.9%.

Table 5.1: ETM-HWP multi-model run results, KPI's represent the mean during the run

Disaggregation method	KPI: Cashflow (€/h)	KPI: TMK (€/kWh)
Baseline	90531	16.23
Weibull	89094	16.14
Triangular	90165	16.10
Rolling average	90510	16.40

5.2. Case 1 disaggregation methods validation

This section aims to validate the disaggregation methods used in the case 1 coupling. Validation of the models used is not the goal. Models used in the couplings are assumed valid and validation of these models is therefore out of scope. The disaggregation methods that facilitate the coupling are tested on their impact on model performance and on their resistance to the typical issues multi-resolution modelling exemplified in table 3.2.

5.2.1. Effect on multi-model performance

The sensitivity analysis performed with the original HWP model in 4.2.5 was repeated for each different disaggregation method of the case 1 coupling. Results of the triangular and weibull distributions are shown below in figures 5.6 and 5.7. The remaining results can be viewed in appendix A.5.

Cashflow	Power +	Power o	Power -
Eprice+	1.2117	1.1025	0.9931
Epriceo	1.0990	1.0000	0.9009
Eprice-	0.9863	0.8975	0.8087

TMK	Power +	Power o	Power -
Eprice +	1.0012	1.0000	0.9982
Eprice o	1.0012	1.0000	0.9982
Eprice -	1.0012	1.0000	0.9982

Figure 5.6: case 1 input sensitivity analysis, triangular disagg

Cashflow	Power +	Power o	Power -
Eprice+	1.2120	1.1028	0.9936
Epriceo	1.0989	1.0000	0.9010
Eprice-	0.9858	0.8972	0.8085

TMK	Power +	Power o	Power -
Eprice +	1.0018	1.0000	0.9981
Eprice o	1.0018	1.0000	0.9981
Eprice -	1.0018	1.0000	0.9981

Figure 5.7: case 1 input sensitivity analysis, weibull disagg

Similarly to the KPI's in table 5.1, figures 5.6 and 5.7 show that the various disaggregation methods have the same effect on sensitivity. When comparing to the original sensitivity analysis done in 4.2.5, the model seems to behave similarly in terms of sensitivity as well. Judging by the results in this section and in section 5.2 it can be concluded that in terms of multi-model performance and sensitivity, there is no meaningful difference between the use of the disaggregation techniques.

5.2.2. Error margins in disaggregation methods

Equation 4.5 specified that a margin of error, E_{thresh}, could be needed to reach a successful disaggregation. Three of the four disaggregation methods use some sort of distribution to draw hourly values (the baseline disagg merely copies the daily values). Given this use of distributions in combination with threshold-based constraints to create a sufficient disaggregation, the error threshold goes hand in hand with the allowed number of iterations that the functions are able to perform. However, the distributions used seemed to have limits to the tightness of the error threshold. At this limit the disaggregation function is unable to create a successful disaggregation that fits the constraints, even when given a large number of iterations. Table 5.2 shows the lowest used error thresholds, as per equation 4.5, alongside the amount of iterations required to create a successful disaggregation.

Table 5.2: ETM-HWP error thresholds and needed iterations

Disaggregation method	Error threshold	Iterations
Baseline	±0%	1
Weibull	±55%	20,000
Triangular	±2.5%	200,000
Rolling average	±0.5%	100

As previously stated, the error threshold can be viewed as the induced translation error during disaggregation. The baseline and rolling-average disaggregations seem to have little to no problem disaggregating within tight error margins at a low iteration count. For the baseline the reason is clear: it copies the day value directly to each hour, therefore there is no error. The rolling average method uses a uniform distribution as seen in equation 4.1. Uniform distributions favor neither a bias upwards or downwards of the mean, therefore making adhering to a mean check easier than it would be for non-uniform distributions. The triangular and weibull disaggregations in table 5.2 confirm this suspicion further. While the triangular is still able to successfully disaggregate at a low error threshold, it requires several orders of magnitude more iterations to do so. This could be due to the bias upwards or downwards from the mean that a triangular distribution can have. The weibull method is able to successfully create a disaggregate. However, it requires an error threshold that is 20 times higher than the second highest alongside a considerable amount of iterations. Judging by table 5.2 it can be stated that the weibull method adheres the poorest to the consistency equation 4.5, while the rolling average method is the best distribution-based disaggregation.

Wind data gathering in physical systems is prone to errors as well. Error sizes in physical wind data gathering could serve as a ballpark estimation for an acceptable error threshold for disaggregation functions. The error in recording wind data is around 0.3m/s without any upwards or downwards bias (Geertsema & Van den Brink, 2014). Given the average windspeed from the ETM dataset is around 7.3m/s, the measuring error of this data in the real system would equal 4.1%. The weibull and triangular method fall below this threshold, which can be interpreted as the disaggregation error being smaller than the measuring accuracy limit in the physical system.

5.2.3. Resistance to thrashing

Thrashing occurs when disaggregation and aggregation happens frequently in succession. While the case 1 coupling is currently not setup in a way that would allow this to occur, possible future expansion of the coupling (or re-use of the disaggregation functions) can still warrant an investigation into the thrashing resistance of the functions used. To check the resistance to thrashing an aggregate value, a mean wind speed, from one day was disaggregated an then aggregated frequently to explore any biases induced through thrashing. After the tests were completed, the triangular distribution was the only method to result in a structural deviation from the original value. A technical explanation, alongside the figures displaying all thrashing tests can be found in appendix A.4.

5.2.4. Static and dynamic consistency costs

Static consistency cost

The disaggregation functions used all roughly have the same cost in terms of time needed to code. However, the weibull disaggregation requires more knowledge of the system being studied. Translating that knowledge effectively into a disaggregation functions also required a time cost. The disaggregation functions were also coded in such a way that the same consistency checking function could be used on each one. This reduced the time cost of making separate consistency checking functions considerably.

The case 1 coupling requires manual extraction of data from the ETM and subsequent manual insertion of data into the disaggregation function and into the HWP model. This process could be automated at an increased one-time static cost. It is however due to this time cost that this automation was not included in this thesis, in favor of allocating more time to the disaggregation functions themselves.

Dynamic consistency cost

Section 5.2.2 described how error margins and needed iterations go hand in hand with each other. Since more iterations requires more runtime, dynamic consistency cost is also coupled to the error threshold. A distribution is able to generate almost any set of datapoints, given that the desired data is within its pdf and (most importantly) it is given enough iterations. The limits of each disaggregation method were given in table 5.2. For these limits a maximum of one minute was given to complete the disaggregation. The baseline disaggregation was significantly faster that all other methods, since it does not have to draw from distributions or iterate to pass consistency checks.

The size of the data impacted the dynamic consistency cost as well, due to the larger amount of iterations it had to run to achieve a disaggregation that passed the consistency check on each individual day. Especially the weibull and triangular function are impacted by the dataset size. These functions required more iterations to successfully disaggregate each day-value. If the dataset had been smaller runtime is expected to go down as well especially in the weibull and triangular methods. A reduction in dataset size could then free up time that can be used to make error thresholds tighter.

5.3. Case study 1 findings

The case 1 coupling is somewhat lob-sided towards the ETM. The ETM vastly exceeds the HWP in size, both in data and calculations. However, the coupling conceptually makes sense. The ETM is an exploratory model that is being used as an information source for a non-exploratory model. A large conceptual and resolution-based difference lays in how both models handle time. The ETM has no representation of time. It jumps directly to the future target state. The HWP which runs on 60-minute intervals therefore has a much higher time resolution. The ETM does supply some sort of time disaggregated data, such as daily wind speeds and hourly electricity prices. While not a perfect match, these values could be converted to fit the HWP.

Attention should also be given to the different modelling principles that the models are based on. The HWP is a linear optimisation model, while the ETM uses network balancing. As seen in 4.4, the to be coupled inputs have very different effects on model behaviour. While both the inputs that the HWP receives from the ETM are able to affect the aggregated output 'cashflow', the input 'TMK' is only variant to constraint-based inputs. Therefore the constraint-based wind speed input could be labelled as having a greater effect on the coupling as a whole and should be given extra attention.

After constructing and analysing all four disaggregation methods to create the case 1 coupling, the methods can be compared using criteria drawn from the coupling audit. From this comparison, a choice can be made as to which method would be best suited to use in the final version of the case 1 coupling.

- **Fit to original data (pdf)**
  The triangular method has the best fit to the original pdf of the data, out of the three distribution-based methods. However, the rolling average and weibull cannot be counted as less reliable in this aspect, since none of the three methods passed a Kolmogorov-Smirnov test.
- **Verification of disaggregation samples**
  The weibull and the triangular methods are both able to capture the broadness of hourly data indicated by the given daily maxima and minima. The baseline and rolling average methods fall short in this aspect, since they revolve to heavily around the mean.
- **Coupling result and effect on performance**
  All disaggregation methods produced similar results on both model KPI's. Model sensitivity was also largely unaffected by changing disaggregation method.
- **Error margins**
  The rolling average and triangular methods were able to successfully disaggregate with a relatively low error threshold. The weibull method falls short in this aspect as it needed a considerably higher threshold in order to complete the disaggregation. The baseline disaggregation outperformed all distribution based methods by not needing any allowed error or iterations.
- **Resistance to thrashing**
  Out of the four methods tested, the triangular distribution was the only one to show a structural drift away from the original data after thrashing. The cause seems to be inherent to the functioning of the distribution itself.
- **Consistency cost**
  The triangular and weibull distributions required a longer runtime, either due to increased computational complexity or increased iterations needed. While the rolling average method was around twice as fast as the triangular and weibull, it was vastly outperformed by the baseline disaggregation which completes almost instantly. All methods are expected to increase in runtime if the size of the aggregate dataset were to increase.

When looking at the factors presented above, the baseline disaggregation method is quite appealing. The baseline method works in a simple manner, adheres to consistency the best, runs the fastest and does not generate different multi-model outputs than more complex disaggregation methods. However it coarseness the entire HWP model down. This could mean that this disaggregation method is best. On the other hand, if a lower resolution input generates the same outcome, one can also wonder whether the original model is not simply operating at a resolution level that is higher than necessary.

5.3.1. Case 1 multi-model run insights for the energy transition

The case studies in this paper mainly serve to iteratively improve the auditing method. However, a successful coupling of two energy models was still made. Therefore, the results of the multi-model run itself can be viewed as a prove of concept of the added value of multi-modelling for the energy transition.

Case 1 presents a bright outlook for the profitability of windfarms in the future. The multi-model is set to run a scenario for 2050. The results are displayed in table 5.3. The cashflow KPI rises by 76.3%, which could be attributed to the higher energy prizes supplied by the ETM. This multi-model shows that there is proper potential for increased revenue for windfarms in the future, especially when keeping in mind that there are no fuel costs associated with wind power that might increase in price in the future. Do note that this multi-model does assume that the general structure of the energy market remains the same (as modelled in the ETM). The second KPI (TMK), which can be interpreted as the added value of the internal hydrogen buffer of the windfarm, rises slightly by 2.66%. While not a move as significant as the Cashflow KPI, there is still a small uptrend. The original HWP model concluded that the added value of a hydrogen buffer for a windfarm would be advantageous (Boereboom, 2020). The driving factor behind the added value of the hydrogen buffer is the imbalance price, which is not currently included in the coupling. Further research on the development of imbalance prices can help to increase the usefulness of the case 1 multi-model.

Table 5.3: Case 1 multi-model run comparison

Run	**KPI: Cashflow (€/h)**	**KPI: TMK (€cent/kWh)**
Original HWP solo model (run setup year: 2016)	51097	15.8
ETM-HWP multi-model (run setup year: 2050)	90075 (+76.3%)	16.22 (+2.66%)

5.4. Case 2 coupling

The case 2 coupling requires two types of disaggregations: a spatial disaggregation of population over municipalities and a temporal disaggregation of price data from hours to 15-minute intervals.

5.4.1. Population spatial disaggregation

To distribute the national population data to municipality level, a reference distribution is needed. There are three ways of referencing the disaggregation that stand out:

1. **A baseline of disaggregation**
   As the amount of municipalities is known, a baseline of disaggregation would be to evenly distribute population across all municipalities.

2. **Total population based disaggregation**
   Figure 4.13 in section 4.5.3 shows the distribution of population in the Netherlands per municipality. It immediately stands out that the Dutch population is highly concentrated in a few municipalities. Using this distribution as a reference for disaggregation would extend this concentration to the distribution of EV-agents.

3. **Age-specific population disaggregation**
   While EV's are gaining in popularity, they remain a luxury available for a select subsection of the population. In particular middle-aged men are prime buyers of electric vehicles (Fuels institute, 2021). CBS (2022) provides population distribution statistics for middle-aged adults (between 45-64), which can be used to divide EV-agents among municipalities as well.

5.4.2. Price temporal disaggregation

Given the information present on electricity prices, there are three ways of disaggregation that stand out:

1. **A baseline of disaggregation**
   Section 4.3.5 shows a coarse means of disaggregation, where the mean price per hour equated to the price per quarter-hour. This disaggregation method can function as a baseline to compare other disaggregation efforts.

2. **Normal-based disaggregation**
   Electricity prices tend to follow a normal distribution (Zhou et al., 2009). A normal distribution fitted to the ETM data could therefore be a potential method of disaggregation.

3. **Rolling-average**
   A rolling average method is able to capture general trends in the data in a quick and repeatable way. If distribution-based options are not viable, the rolling average method can be a proper solution as well

Section 4.5.3 states that a normal distribution is set as the standard for predicting electricity prices. However, it also exposes that the discrete price values provided by the ETM do not fit a normal distribution well. The lowest and highest values in the ETM price data are 0 and 117.48. If a disaggregation function were to be based on the normal distribution from figure B.1, it would have to be able to draw around the max and the min values within an error threshold. Unfortunately, the chances that the normal distribution of figure B.1 draws around 0 or 117.48 are 1.5e-22 and 1.2e-50. These chances are so low that a normal-based disaggregation function would either need an unrealistic amount of time to iterate or a disproportionately broad error threshold. Neither of these are desirable. Therefore the normal-based disaggregation method is discarded due to its infeasibility.

5.4.3. Disaggregation techniques verification

The proposed population distributions have been normalised and multiplied by the national total of EVs to produce an EV-agent count per municipality. The accompanying distribution graphics can be seen in appendix B.6.

A sample of the rolling-average-based price disaggregation is presented in figure 5.8. It is clear that an increase in window size moves the disaggregate further away from the original values. However, as stated in 4.5.3, this is not necessarily a bad thing. The moving averages do follow the directional difference in the original data. The window size also affects this, for example resulting in the 12 and 16 hour moving average opposing the direction of the 4- and 8 hour moving averages around t = 88 in figure 5.8. The window-based variation is to be expected, as the larger windows capture more information and thus produce a smoother curve than the smaller windows. While shorter windows may be able to more accurately follow the direction (see section 5.5.2), longer windows produce fewer periods of prices being the same. As real electricity prices are rarely the same for two 15-minute intervals (Zhou et al., 2009), this increased diversity is favourable in terms of realism.

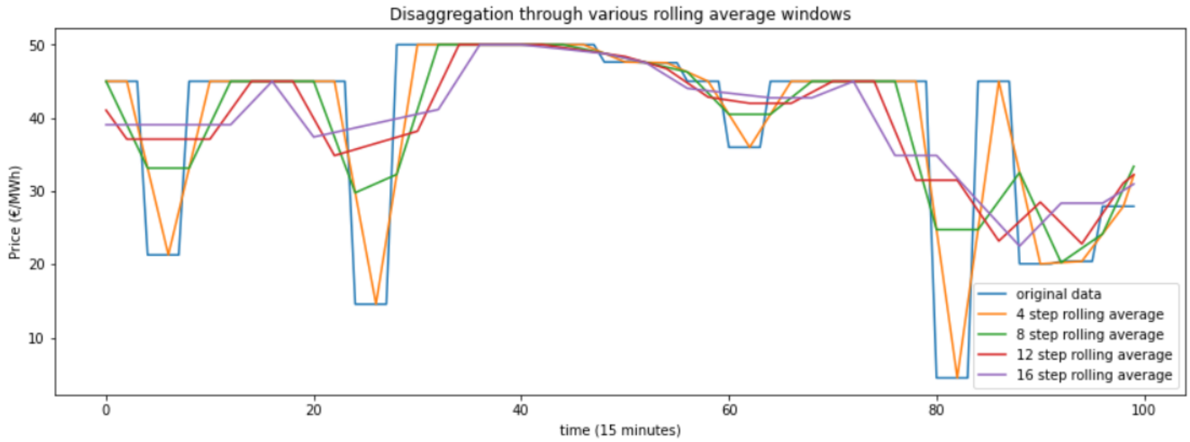

Figure 5.8: Rolling-average-based disaggregation sample using various window sizes

5.4.4. Coupling run results

Table 5.4 shows the KPI results from the case 2 coupling, using the various disaggregation methods. Only one dimension (space or time) was varied per run in order to clearly show the effect of each method. Given that the original EVM used the total population of The Netherlands to allocate EV-agents to municipalities (Boendermaker et al., 2022), the corresponding method 'Total population' can be seen as a baseline reference point.

The EV population data from the ETM provided a computational challenge, because the total agent count would surpass 10 million agents ($\pm 12.3M$ agents). Running the multi-model with this large agentset takes 48+ hours per run on the available machinery. The choice was therefore made to reduce the agent count to 25,000 agents during the exploratory runs. A comparison of output KPI's using the full 12.3M agentset can be found in section 5.6.1.

Table 5.4: ETM-EVM multi-model run results

Disaggregation dimension	Disaggregation method	KPI: Total power demand	KPI: Vehicle to grid capacity
Space	*Baseline*	4248.7	55185.1
	Total population	4101.3	53915.2
	Age-specific	4266.9	55206.6
Time	*Baseline*	4130.9	61995.9
	Rolling average 4	4111.5	54772.5
	Rolling average 8	4111.0	54640.4
	Rolling average 12	4111.0	54640.4
	Rolling average 16	4105.0	55116.1

The largest deviation from the 'total population' reference happens during the 'Baseline' run. This method causes a ±15% deviation in the KPI vehicle to grid capacity. The rolling average methods significantly reduce this deviation to around 1.4%, which can be seen as relatively indistinguishable. The deviation of the Baseline time disaggregation method can be explained by earlier findings in the performed audits that show that the price differentials between timesteps is important for agent behaviour. The baseline does not translate the intra-day price differential at all, which results in the change in outputs shown in table 5.4.

5.5. Case 2 disaggregation methods validation

5.5.1. Effect on multi-model performance

Same sensitivity analysis as from section 4.3.5 yielded similar results when done using the various disaggregation methods to generate the input data. The full results of these sensitivity tests can be seen in appendix B.7.

The lack of variety in the sensitivity analyses conceals the differences that occur in the model on a smaller level. The output KPI's of the EVM are on a national level, aggregated from the municipality-level model. While the national results may be the same across the various disaggregation techniques, large variations occur on the municipality level. For instance, The municipality 'Amsterdam' receives vastly different amounts of EV's between the disaggregation techniques. Under 'baseline' spatial disaggregation Amsterdam receives 0.28% of the EV-agents, under 'age specific' 0.23%, but under the 'total population disagg it receives 5.00% of EV-agents. Figure 5.9 (graph of VTG KPI in appendix B.8) shows that there are in fact large differences present between disaggregation methods, but that some only become present on the higher resolution level.

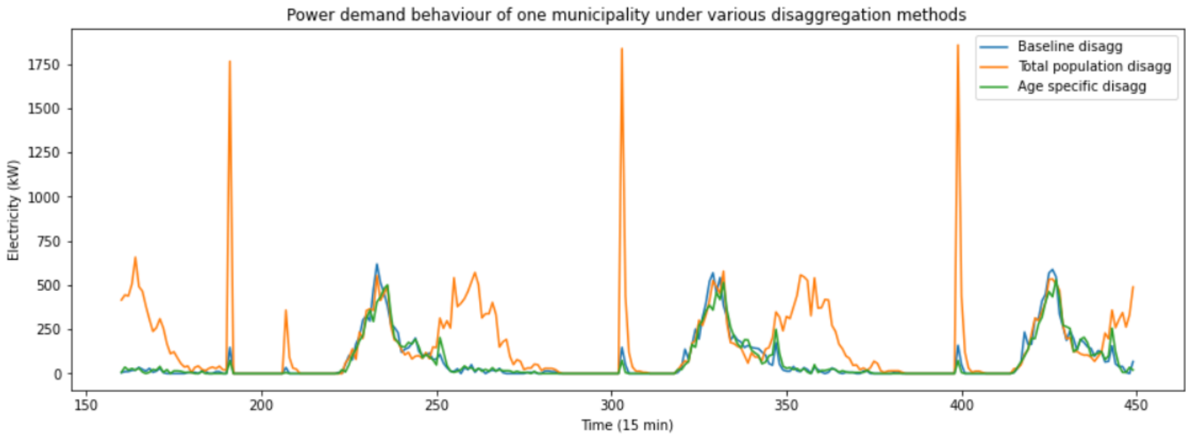

Figure 5.9: Power demand of municipality 'Amsterdam'

5.5.2. Error margins in disaggregation methods

Using equation B.1 and B.2, the rolling averages can be checked for directional similarity. The results of this test can be seen in figure 5.10. Interestingly, there seem to be several levels where the similarity percentage plateaus. The 4-hour and 8-hour rolling average therefore roughly have the same similarity percentage of +-70%. One explanation for these plateaus is that the original data is in hours divided into 4 15-minute intervals. Therefore, if the rolling window increases by 4, the disaggregation method will incorporate more differing values. The plateaus present in figure 5.10 do not follow this pattern of 4 perfectly however.

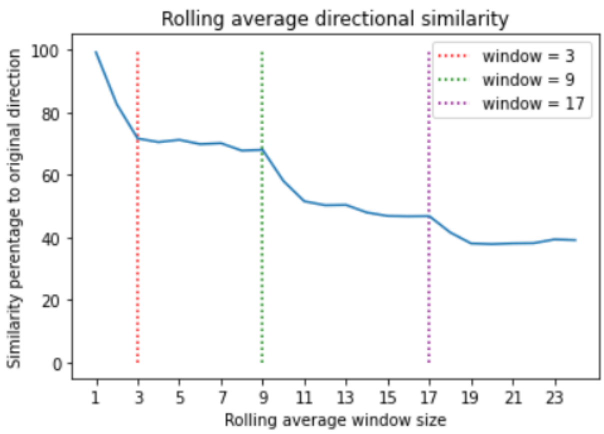

Figure 5.10: case 2 rolling average derivative similarity

Table 5.5 gives an oversight of the specific error thresholds and accompanying iterations. Since there are no distributions used in spatial and temporal disaggregations, only one iteration is needed to create a successful disaggregate dataset. Error thresholds in the spatial disaggregation are mainly caused by rounding errors during the disaggregation. Since the agent count per municipality can only be an integer, any floats resulting from the disaggregation are rounded off. This error fortunately remains small enough to remain insignificant.

Table 5.5: ETM-EVM error thresholds and needed iterations

Disaggregation dimension	Disaggregation method	Error threshold	Iterations
Space	*Baseline*	±0.003%	1
	Total population	±0.001%	1
	Age-specific	±0.002%	1
Time	*Baseline*	0%	1
	Rolling average 4	±29.5%	1
	Rolling average 8	±32.2%	1
	Rolling average 12	±49.7%	1
	Rolling average 16	±53.3%	1

5.5.3. Resistance to thrashing

Since prices are disaggregated using a rolling average function only, there is no randomness used to create the disaggregate values. Therefore, an accompanying aggregation function could be made that perfectly reverses the disaggregation. Such a reversal function was created, which resulted in the ability to nearly perfectly reverse a disaggregation effort. Therefore, all rolling average-based time aggregations can be considered resistant to thrashing. A full technical breakdown of the rolling-average-based thrashing tests can be found in appendix B.2. The accompanying figures can be found in appendix B.3.

The space disaggregation distributes an integer number of EVs accross municipalities. This proces can be reversed easily by summing all the EVs from the municipalities back up together. As a result, all spatial disaggregations were also found to be resistant to thrashing. Any errors that did occur were caused by rounding errors due to the requirement that the amount of EVs in a given municipality must be an integer. Fortunately, this error was not able to create any meaningful drift during thrashing. All figures for the space thrashing tests can be found in appendix B.4.

5.5.4. Static and dynamic consistency cost

Static consistency cost

The functions needed for the space and time disaggregation functions all require roughly the same time to code. The main cost in terms of time comes from the space disaggregation, namely finding and processing the proper datasets to use as reference for population per municipality. These datasets also become dated after use. In contrast to other disaggregation functions, which are mainly mathematical operations, the spatial disaggregations revolve around these distribution datasets. Therefore it could be needed to update these datasets over the years. Technically this updating would fall under dynamic consistency, since it is recurring. However given the infrequency and the nature of this operation, it can better be viewed as maintenance under static consistency.

Dynamic consistency cost

The runtime for all disaggregation methods used was low, especially when compared to case study 1. This was to be expected. No distributions were used, removing the need to iterate towards a successful disaggregation and thus significantly reducing runtime. However, estimated high dynamic consistency costs in section 5.4.2 were among the reasons to outright exclude the normal distribution method for further use, even though literature suggested it would fit well. Would the data from the ETM have been more favourable towards distributions, the normal distribution might not have been excluded. The functions that remained were therefore not distribution based and required much less time to complete.

5.6. Case study 2 findings

Similarly to case study 1, the ETM exceeds the EVM in size. However due to its size, the select few pieces of information that the EVM requires are of a lower resolution. Time resolution differs between the two models, and careful attention should be payed to how temporal data such as electricity prices are disaggregated. The EVM is an agent-based model that heavily relies on the behaviour of the agents to create model results. These agents perceive time per tick, resembling 15-minutes. It is of no influence to the agents in the EVM what the general price of electricity is, whether the average price is €1/kWh or €1000/kWh. Agents base their price-related behaviour solely on the fluctuations over time. It is therefore the aim of the temporal disaggregation functions to translate the fluctuations between prices, not the price levels themselves.

The EVM is a national-scale model that runs on a municipality level. Surprisingly, the ETM is able to run on this level as well. However, the national-scale ETM model is not an aggregation of the models at municipality level. This is mainly due to data constraints such as that municipal datasets, when added up, do not match national-level statistics (Quintel, 2022c). Therefore the national data of the ETM is used and must be disaggregated to spatial level. Logic specific constraints then become very important in this spatial coupling, because a distributive key is needed to give a realistic distribution of agents. Mathematical distributions can be used, but these would likely give agent distributions that would be difficult to call valid.

After constructing and analysing all temporal and spatial disaggregation methods to create the case 2 coupling, the methods can be compared. From this comparison, a choice can be made as to which method would be best suited to use in the final version of the case 1 coupling.

- **Fit to original data**
  Spatial: The EVM originally used the total population of the Netherlands to distribute agents with. Logically the disaggregation function based on the total population the this original method the best. Both the baseline and age-specific distributions spread the agent population out more over all municipalities, while the total population method concentrates most of the agents in municipalities such as Amsterdam and Rotterdam.
  Temporal: From an initial visual inspection, the 4-step rolling average method resembles the original baseline data the most.

- **Verification of disaggregation samples**
  Temporal: While the 4-step rolling average is able to follow the baseline the most, it also creates larger areas where electricity prices are constant. Section 4.3.5 shows that having large sections where the price is the same affects model behaviour. This is undesirable, given that electricity prices in the physical system rarely remain constant. Higher rolling windows reduce these periods of constant prices, at the cost of tracking the original price levels less well.

- **Coupling result and effect on performance**
  All disaggregation methods produced similar results on both model KPI's. While model sensitivity was indifferent to changes in disaggregation method on the national scale KPI's, municipality level behaviour is significantly affected by changes in spatial disaggregation methods.

- **Error margins**
  Spatial: The error margins are negligibly low. The small error that does occur stems from rounding errors of the float numbers generated by the disaggregation, caused by the fact that the resulting agent-count per municipality must be an integer.

  Temporal: The error margin for the temporal disaggregation measures the level to which the disaggregate dataset matches the direction of the original. The 8-step rolling average strikes a good balance between having a relatively good directional similarity, while reducing the amount of constant-price areas when compared to the 4-step window.

- **Resistance to thrashing**
  All functions showed excellent resistance to thrashing. It is important to note that the rolling average methods requires the storage of some 'junk information' to be able to aggregate back to the original data.

- **Consistency cost**
  The consistency cost was kept low in all disaggregation methods, namely due to the exclusion of any techniques based on probability distributions. Because of this, iteration is not needed and aggregation can be done much more accurately.

The described facets of each technique above, in combination with the more detailed analyses of chapters 4 and 5, suggest that the 'Total population' spatial disaggregation and the '8-step rolling average' temporal disaggregation are best suited to facilitate the coupling of case study 2. The 8-step rolling average strikes a fine balance between the key things that a this particular coupling needs, namely translation of direction combined with reduced invariance in electricity prices (while adhering to acceptable consistency maintenance). The spatial disaggregation choice is ultimately based on the knowledge of the physical system for deciding which distribution key can be used to allocate the EV-agents. Based on the national-level output KPI's, any of the proposed spatial disaggregations would suffice. However, the question becomes if there will be additional coupling done in the future on municipality level that works with the ETM-EVM coupling of case study 2. In this case, the choice of spatial distribution key becomes more significant. It is therefore advisable that such spatial disaggregation functions be made in a way that different distribution keys are easily interchangeable. This was the case in this case study, which resulted in both ease of use and reduced static consistency cost.

5.6.1. Case 2 multi-model run insights for the energy transition

The multi-model of case 2 provides some interesting insights for the energy transition. The 2050 scenario used in the ETM outputs a considerable increase in EVs, going from 174,000 to 12.3 million EV agents (+7069%). An increase on the same scale can be seen in the output KPI's in table 4.6. Linear growth could reasonably be expected. An increase in vehicles was not expected to lead to exponential growth of power demand for example. What the multi-model does show us is that the modelled 2050 power demand and VTG capacity are prone to large peaks. For example, power demand momentarily peaks to 69.1GW during the run. Such peaks are far beyond the current power infrastructure to handle. However, these peaks occur due to the accumulation of EVs smart charging at the cheapest points in time. The multi-model points to the fact that such an accumulation of smart charging is not desirable due to the overload it causes on the system. A conclusion from the multi-model in case 2 can therefore be that price sensitive smart charging as used in the multi-model causes undesirable peaks in power demand that must be mitigated in order to prevent infrastructure overloads. One way that this could be done is for smart charging software in EVs to strive for not only the cheapest price, but also an even spread of charging demand over time.

Table 5.6: Case 2 multi-model run comparison

Run	**KPI: Total power demand (GW)**	**KPI: Vehicle to grid capacity (GW)**
Original EVM solo model (run setup year: 2021)	0.0299	0.362
ETM-EVM multi-model (run setup year: 2050)	2.129 (+7097%)	25.746 (+7114%)

6 Discussion

The main research question of this thesis was as follows:

How can issues that arise when coupling multiple energy models that have different resolutions be resolved effectively?

To answer the main research question, a method of auditing is proposed that creates two lists of key questions (audits). These guided the modeller through the process of coupling the models in the case studies. The lists are created by extracting key issues from literature on multi-resolution modelling, multi-modelling in general and consistency maintenance in multi-modelling. The audits are phrased in a way to generate an overview of both the individual models that are to be coupled and the coupling itself. Based on the audits and the case studies done in the previous chapters, this discussion aims to reflect on the effects and limitations of the auditing methods (sub-question 5, see 1.2). The discussion is organised in exploring the effects, downsides and unexplored facets of the auditing method as a means of answering the research question. In addition, attention will also be given to when this method would be fit to use for anyone else working on multi-resolution multi-modelling.

Effects

The audit questions were created by examining the problems that other modellers encountered in earlier multi-resolution multi-modelling efforts. Recurring problems and solutions in these papers were included to be part of the audits. The two audits create an overview of typical issues that can occur in multi-resolution multi-modelling. This was the aim set in sub-question 2.

The completeness of the audits was tested with two case studies of hypothetical couplings. In the two case studies, following the auditing method resulted in the successful recognition of coupling issues present (sub-question 3, see 1.2). A clear benefit during the case studies was the standardised way that the audits lead to information generation. Laying two model audits next to each other pointed out inconsistencies and oddities in a natural way. An example of this was the recognition of the EVM's sensitivity to price change (not price level) in case study 2. However, the ETM provided prices in discrete levels, that were difficult to disaggregate using probability distributions. This difference lead to the consistency maintenance method being changed from an equal average rule to a preservation of price change. While filling in the model audits there was a natural familiarisation that occurred. Not only were coupling issues recognised, but it also became clearer what the 'coupling' in itself really meant.

The coupling audits were more practical and set the base for what the A/D functions needed to accomplish. Using the coupling audits' questions on consistency, different disaggregation types could be compared on several different facets. Which of these facets is most important depends on the specific case it is in. An example of this in case study 1 is the thrashing tests done on the various wind speed disaggregation methods. It became clear that the triangular distribution showed a structural bias away from the original values during thrashing, while the other disaggregation methods did not. The issue was not deemed a critical flaw for the disaggregation method, because the coupling in case study 1 was not prone to frequent A/D. Nevertheless, the insight gained from exploring the triangular distribution in this way could be of use in future coupling efforts.

Insights of the model audits combined with the A/D oversight of the coupling audits resulted in successful couplings in both case studies (sub-question 4, see 1.2). The two case studies were able to create ample room for the testing and iterative improvement of the audits. An example of such an improvement was in case 2, where it became clear that sensitivity analyses on different price disaggregation methods did not show any difference. Previous insights of the model audit did indicate that the EVM would be sensitive to price change. Given that A/D was able to alter the change in the price patterns, a question was added in the model audit that inquired about sensitivity to change in input *resolution*, not just input level variation. By altering the resolution of price to be more coarse (more subsequent timesteps with the same price) the model did show different behaviour. This analysis lead to the insight that coarseness in terms of long periods of constant prices was undesirable from a disaggregation method, and should be avoided if possible.

Limitations

While following the proposed audit method lead to the creating of successful couplings in the two case studies, there can be no guarantee that the audits as they are currently would be sufficient in finding resolution-based issues in other coupling cases. In fact, there is also no guaranty that all possible resolution-based issues in the two case studies were detected. A recommendation that then naturally comes to mind for mitigating this limitation in the auditing lists is to expand them to be more thorough. This sentiment can be best phrased by the famous Wikipedia disclaimer *"This list is incomplete. You can help by expanding it"*. It is not the intention of the thesis to state that the lists proposed are complete and able to uncover all resolution-based issues in energy multi-models. These lists are an iterative product of literature research and trial and error through case studies. One relatively hidden benefit of the lists is that they are relatively short and concrete. This is an advantage that should not be overlooked. The overview that the model audits provide is concise, which has proven to be of use during the case studies to keep oversight and create insight. One could theoretically expand these lists indefinitely to uncover more and more about a models' issues and inner workings. However, at one point or another the law of diminishing returns would start to kick in. If the only models that are to be coupled are models that are 100% understood, it would be a major hurdle for coupling models not previously meant for coupling.

Another limitation of the audit method is that while they do propose checking for consistency in A/D methods, they do not specify any hard limits on what constitutes 'good' consistency. What is an acceptable threshold? When are semantics sufficiently aligned? The question of 'what is best' can be answered partially with literature, knowledge of system and logic specific constraints and the overview of 'if it makes sense' from the audits. An example of this is case study 1, where the consistency thresholds of the wind speed disaggregations were compared to measuring errors in the physical system. Even though, the problem remains that there is still no hard limit for a 'good' consistency threshold. From literature and the case studies it was found that 'good' consistency depends on the needs of the coupling effort being done. It is likely that during a coupling effort, there is no one disaggregation method that excels in all consistency checks (if there is, that is great!). What the audits do accomplish is to give a variation of checks that can be used to compare how various A/D methods perform in various categories of consistency maintenance. It is then up to the modeller to inform the stakeholders in the multi-modelling effort clearly on what a satisficing solution could be and what the associated risks of not e.g. holding to a stricter consistency standard would lead to. for example, in the case studies done, the negative effects of thrashing were limited, thus giving less weight to a negative result in this category. Such a finding may drive the choice to favour other values like speed over thrashing consistency.

As a final limitation, sub-question 4 asks what methods can be used to alleviate the issues present. The audits, while guiding towards A/D functions, do not mentioned any specific disaggregation methods A, B or C that a modeller should use when the coupling issues resemble X, Y or Z. However, the case studies in the thesis do show the process how the disaggregation methods were chosen based on the information gathered from the audits. Given the pluriformity of models and the even larger set of possible coupling configurations, it would be almost impossible to generate specific instructions for every possible A/D effort. This is the reason why the decision was made to forfeit this specificity in the audits in favour of a more general method of guiding towards proper A/D methods.

Unexplored facets

Some facets of finding typical issues in multi-resolution multi-models remain unexplored in this thesis. Three of the unexplored avenues are listed below. These currently unexplored subjects could serve as continuation points for future research.

Both case studies included models with a focus on energy and energy systems. This is partially due to the availability of the used models at the time of writing, alongside the writers interest and knowledge in energy systems and energy modelling. During the case studies it became very clear that domain specific knowledge is needed to complete the couplings. This knowledge has proven vital in the coupling process in areas such as semantics comparison, the understanding of appropriate distributions to use for A/D and the setting of logic specific constraints for the A/D functions. While the implementation of the audits in the two cases required energy-domain specific knowledge, the audits themselves are purposefully not geared towards energy in particular. The audits were created with an aim of more general use in mind. However, whether the audits as they are currently are able to accomplish the coupling of models in other domains remains unknown. It could be that certain domains include specific issues in A/D that would require additions to the questions in either the model or the coupling audit. Further exploration of the use of the audits in other domains can help in understanding if the audits are capable of more generally helping in the recognition and alleviation of typical issues in multi-resolution multi-modelling.

Another unexplored subject revolves around the effect of increasing the complexity of the coupling case. Increased complexity could entail a tighter coupling of frequent two-way communication between two models, or the coupling of more than 2 models in a multi-resolution multi-model. The audits were created iteratively from the two case studies. However, both case studies used a one-way coupling between two models. It is unknown how increased case complexity would affect the audits ability to recognise and help alleviate resolution-based issues. The literature hints at the importance of tighter consistency thresholds, especially when the couplings become increasingly complex. For instance, a multi-model consisting of 5 models may feature multiple ways for information from model A to reach model B (either direct or indirect). This has overlap with the problem described in figure 2.3. Although these types of avenues remains unexplored, the literature and audits present in this thesis do already contain some knowledge aimed at recognising resolution-based coupling issues in more complex or dynamic couplings. Another example of this is the included inquiry about thrashing resistance. This is a result of the audits being created from literature concerning previous multi-resolution multi-modelling efforts, of which some described models were more complex and dynamic than the case studies in this thesis.

A quite different unknown facet of the research is the possible biases in the composition of the audits themselves. The audits were drafted with the intent of being unbiased and objective inquiries about the facets of models that could lead to resolution-based coupling issues. However, the audits were drafted by a human and remain a product of iterative improvement using two specific case studies. It is therefore not unreasonable to assume that some bias in the audits can be present. An example of bias could be that certain issues are given more or less attention, based on the knowledge and understanding of topics of the writer. Another bias could be that linguistically the audit questions are phrased in such a way that guide the reader to certain flaws that were specifically present in the case studies used. If a specific issue was particularly present in one of the case studies, a question inquiring about this issue might have been added to specifically find this issue in the future, making the list more over-fitted and biased to solving the presented case studies and less focused on the general recognition of typical issues. The presence of such biases are hard to uncover, partially because it has proven difficult for the writer to spot one's own biases in the middle of the thesis process. The presence of biases and the extent to which they influence the usefulness of the audits therefore remains unknown.

Applicability

The contribution that this thesis can offer the modelling community lies in the applicability of the audits in other coupling efforts. Therefore it is useful to reflect on what type of coupling settings the audits may or may not be applicable on. As a further clarification, the audit method that may be applicable for use in other modelling efforts refers to the process described in figure 6.1. It entails (for a two-model coupling) a separate model audit for each model, leading to a coupling audit using both models and finally to the realisation of the coupling itself.

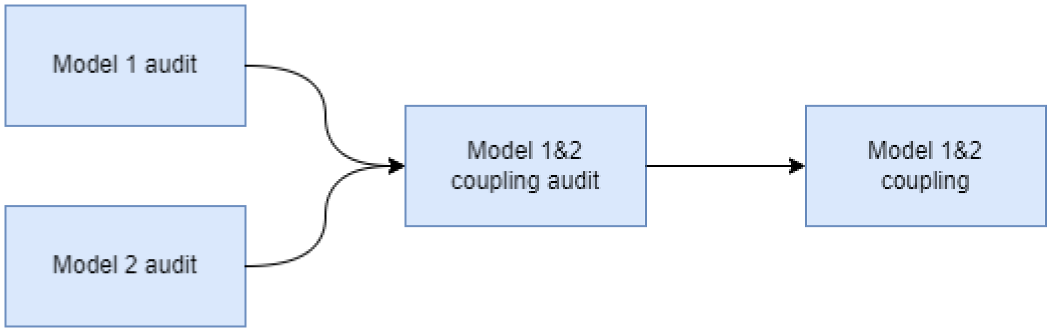

Figure 6.1: Graphic representation of audit method

When looking back on the process of creating the audits and the case studies that came with it, a certain type of coupling efforts stands out as being well suited to use the audits. A modeller could consider using the audit method when faced with:

- Coupling two energy models not previously intended for coupling, which operate on the same scale but not on the same resolution.

Such cases are what the audits were iteratively built on. Thus it can be estimated that the audits as they are currently will likely be best suited to aiding in similar coupling efforts. Multi-resolution model coupling usually has an ad-hoc nature to it. A coupling for a selection of models have to be made, and so it is made. Using the audits standardises the process to some extent, which enables a more structured and comparable approach. The standardised way that the audits guided the two case studies helped in increasing the speed at which couplings could be done as well. Once the first case was completed, the second case took less time to come to a successful coupling. Whether the application of the audits can be further broadened to non-energy models or more complex multi-models is currently uncertain (and listed above).

The audits would be less suited if the modeller in question is faced with the following two coupling challenges:

- A coupling of two (or more) models on differing resolutions *and* on differing scales. Handling issues related to coupling models on different scales was deemed out of scope for this thesis, in order to better focus on resolution-based issues. As a result, the audits are unequipped to help recognise and alleviate issues resulting from scale differences.
- A coupling of two (or more) models which are intended to be coupled according to a strict coupling method, such as an HLA. The audits were specifically designed to aid in the multi-model infrastructure (Multi-model.nl, 2021), which deals with coupling models not originally intended for coupling. Models that are created from the start with the intent of being coupled are better off adhering to the specific coupling guidelines of the multi-model that the model was intended for.

7 Conclusion

In order to aid in the effort to reach the 2030 & 2050 dutch climate goals, a broad understanding of Dutch energy infrastructure and the models concerning it is needed. This thesis is part of a larger project that aims to create a multi-model infrastructure to couple existing models from chain partners into one platform, in order to attain better understanding and create more detailed policy. The problem that this thesis aims to address in the larger project is finding and resolving typical issues that appear when attempting to couple models at varying resolutions, that weren't originally built to be coupled, into the multi-model infrastructure. The main research question of this thesis was therefore as follows:

How can issues that arise when coupling multiple energy models that have different resolutions be resolved effectively?

This thesis proposes the use of two audits, a model audit and a coupling audit, in order to effectively find resolution-based issues and to guide a modeller towards methods of alleviating these issues. The main conclusion of the conducted research is:

The problem described in the main research question can be answered by using a coupling process based on audits, comprised of questions aimed at detecting issues and checking the effectivity of the means to solve the issues.

The proposed audits create a method to explore possible resolution-based issues in a standardised manner for coupling the energy models present in the case studies. The audits are comprised of questions aimed at recognising typical issues found in literature regarding topics of multi-resolution modelling and multi-modelling in general. In order to further improve the audits, two case studies were done which each describe a hypothetical multi-resolution coupling of two energy models. Both case studies used the audits to successfully go through the full process of finding typical issues, resolving them and completing the coupling. From the research done in this paper it can therefore be concluded that the proposed audits were an effective tool to recognise and help alleviate the typical resolution-based coupling issues present in the two documented case studies.

The challenges presented during coupling can be divided into two categories: recognising which issues are present and designing effective methods to alleviate them. From the research done, the second conclusion is:

The recognition of which typical issues are present in a coupling effort is the tougher challenge of the two.

During the case studies, once the (possible) issues present were clear, the actual creation of the disaggregation functions proved less of a challenge. In the case studies the audits provided guidance towards proper aggregation and disaggregation (A/D) functions. While the audits do not give specific recommendations as to which A/D functions to use, the process does facilitate a more directed and standardised selection and evaluation process. The Multi-model.nl stands to benefit from this standardisation. Many couplings will need to be made during the multi-model.nl project. The audits can standardise the way these couplings are researched and developed. This does not only increase repeatability, but e.g. previously performed model audits could also be re-used when expanding the multi-model infrastructure.

The two conducted case studies served to aid in improving the audits during the thesis. While working on the case studies, several findings about the process of coupling multi-resolution multi-models were done as well. The third conclusion is therefore:

The coupling of two models will likely require the construction of an A/D effort specifically tailored to the needs of the models in question.

It is quite unlikely that an A/D effort done to service the coupling of model A and B can be directly implemented to also facilitate the coupling of models C and D. From the two case studies done it seems that there are always subtle differences that need to be taken into consideration when coupling. Some differences might be more easily bridged than others. For instance, case study 1 required a time disaggregation from days to hours. Case study 2, which required a slightly different time disaggregation (hours to quarter-hours) was able to re-use some parts of the disaggregation function of case study 1. However, other differences might be harder to bridge. For example, the same two time disaggregations in case studies 1 and 2 required very different types of consistency checking functions, based on the findings from the audits. Therefore case study 2 was unable to re-use (parts of) the consistency checking function of case study 1. Do note that the absence of re-usable functions does not imply that learnings on the coupling process can't be translated. In fact, the opposite is true. After the process of case study 1 was concluded, executing the same process for case study two could be done much more time-efficiently. The efficiency increase from case study 1 to case study 2 seemed to come from the familiarity gained in the standardised way that the audit facilitate the coupling process.

Additionally, the case studies made clear that:

System expertise is highly advantageous for identifying and understanding possible problems during the auditing and coupling process.

An example of this is semantics comparison. In both case studies, electricity prices were one of the variables exchanged between the models. However, the semantics of what each model understood as 'electricity price' differed in each case study. Knowledge of (in this case) energy systems proved needed to check whether the semantics were comparable with each other. Semantics were not the only area where system specific knowledge was needed. Other examples include the creation of the consistency checking functions in case study 2 (see section 4.5.3) and the identification of the proper distributions to use for wind speed disaggregation (see section 4.4.3). In short, system specific knowledge is needed throughout the coupling process.

Finally, it became clear in the case studies that:

Such couplings that require an A/D effort need a data model in the middle to translate (and possibly store) data to facilitate the coupling.

The case studies show and analyse various ways that the translation of model A to model B can be done. During the case studies it became clear that a significant amount of data modelling is needed in these steps. A modeller coupling two such models as in the case studies must be aware that inserting a data model introduces a third model into the mix. This third (data) model, just like the other two, can be prone to bias which might influence multi-model behaviour. Much attention was therefore payed in the case studies comparing different methods in these data models to expose possible biases that might influence behaviour. Even though, it remains likely that any data model constructed comes with some level of bias which could influence how the coupling behaves.

7.0.1. Further research recommendations

The discussion in chapter 6 describes several limitations and unknowns which can form good starting points for further research. One of these topics is the specification of limits regarding appropriate consistency thresholds and guidelines to determine what constitutes as sufficient semantic overlap. During the case studies in this thesis these specifications were not set. Instead, an exploratory modelling effort was conducted to clearly reflect what the feasible consistency limits of various techniques were to the stakeholders to which the models belong. The choice if a consistency limit was sufficient in the case studies was then left to the preferences of the stakeholders as well. Further research on hard limits of consistency and semantic overlap could aid in testing A/D efforts and clarifying when and why certain consistency standards should be upheld.

In addition, further research could focus on applying the audits in other types of coupling efforts. Application of audits to other fields or application of the audits in a multi-model comprising of more than two models could clarify the broader usability of the audits. Perhaps certain broader applications require adaptations to the audits, be it minor or major. In any case, the value of the audits can be increased if research is conducted which documents the broader applicability of the method.

Finally, this thesis was intended to aid the coupling effort of Multi-model.nl (2021). The audit process can be of use in recognising and alleviating issues for this purpose. However, the coupling effort of Multi-model.nl (2021) is considerably more complex than the performed case studies. Further research can therefore also be dedicated to testing and improving the effectivity of the auditing method on more complex coupling cases. For example, the studied literature hints that subjects like communication consistency maintenance and thrashing resistance become even more important in complex couplings. It is therefore recommended that the literature and the auditing method present in this thesis be used by the modellers of Multi-model.nl as a starting point for developing audits that are better suited to recognise and alleviate issues in more complex cases.

References

Adler, P., Falk, C., Friedler, S., Nix, T., Rybeck, G., Scheidegger, C., Smith, B., & Venkatasubramanian, S. (2018). Auditing black-box models for indirect influence. *Knowledge and Information Systems*, *54*(1), 95–122. https://doi.org/10.1007/s10115-017-1116-3

Adya, A. (1999, March). *Weak consistency: A generalized theory and optimistic implementations for distributed transactions* (Doctoral dissertation). https://pmg.csail.mit.edu/papers/adya-phd.pdf

Andreas. (2019, August 3). *Free image on pixabay - electric car, automobile*. Retrieved January 13, 2022, from https://pixabay.com/photos/electric-car-automobile-carsharing-4381728/

Bankes. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. https://doi.org/10.1287/opre.41.3.435

Bankes. (2002). Tools and techniques for developing policies for complex and uncertain systems - PubMed. https://doi.org/10.1073/pnas.092081399

Bloomberg. (2022). Renewable energy burst sends dutch power prices to lowest ever. *Bloomberg.com*. Retrieved June 23, 2022, from https://www.bloomberg.com/news/articles/2022-04-25/renewable-energy-burst-sends-dutch-power-prices-to-lowest-ever

Boendermaker, F., Boereboom, B., & Hermans, B. (2022, January). *Vehicle to grid capacity in the netherlands*. Delft, Netherlands. https://github.com/bramboereboom/MSc-thesis.git

Boereboom, B. (2020, July 10). *Microwaterstof* (BSc thesis). Delft university of technology. https://github.com/bramboereboom/MSc-thesis.git

Brandmeyer, J. E., & Karimi, H. A. (2000). Coupling methodologies for environmental models. *Environmental Modelling & Software*, *15*(5), 479–488. https://doi.org/10.1016/S1364-8152(00)00027-X

CBS. (2022). *Inwoners per gemeente*. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/inwoners

Chen, J., & Li, X. (2021). Research on key technologies of multi-resolution modeling simulation. *2021 International Conference on Intelligent Computing, Automation and Applications (ICAA)*, 687–693. https://doi.org/10.1109/ICAA53760.2021.00124

Chua, B. Y. W., & Low, M. Y. H. (2009). Predictive algorithms for aggregation and disaggregation in mixed mode simulation. *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 1356–1365. https://doi.org/10.1109/WSC.2009.5429284

Dahmann, J. S., Fujimoto, R. M., & Weatherly, R. M. (1997). The department of defense high level architecture. *Proceedings of the 29th conference on Winter simulation - Atlanta, Georgia, United States*, 142–149. https://doi.org/10.1145/268437.268465

Davis, P., & Bigelow, J. (1998a). *Experiments in multiresolution modelling (MRM)*. RAND. https://www.rand.org/content/dam/rand/pubs/monograph_reports/2007/MR1004.pdf

Davis, P., & Bigelow, J. (1998b). Introduction to multi-resolution modeling (MRM) with an example involving precision fires. *3369*, 14–27. https://doi.org/10.1117/12.319334

Davis, P., & Hillestad, R. (1993). Aggregation, disaggregation, and the challenge of crossing levels of resolution when designing and connecting models. *1993 4th Annual Conference on AI, Simulation and Planning in High Autonomy Systems*, 180–188. https://doi.org/10.1109/AIHAS.1993.410585

Degbelo, A., & Kuhn, W. (2018). Spatial and temporal resolution of geographic information: An observation-based theory. *3*(1), 12. https://doi.org/10.1186/s40965-018-0053-8

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society*, *57*(1), 45–70. https://doi.org/10.1111/j.2517-6161.1995.tb02015.x

DWIA. (2003, September 19). *Describing wind variations: Weibull distribution*. http://xn--drmstrre-64ad.dk/wp-content/wind/miller/windpower%20web/en/tour/wres/weibull.htm

EZK. (2019, June 28). *Klimaatakkoord*. https://www.klimaatakkoord.nl/documenten/publicaties/2019/06/28/klimaatakkoord

Fuels institute. (2021, June). *EV consumer behaviour*. https://www.fuelsinstitute.org/Research/Reports/EV-Consumer-Behavior/EV-Consumer-Behavior-Report.pdf

Fujimoto, R. (2015). Parallel and distributed simulation. *Proceedings of the 2015 Winter Simulation Conference - Atlanta, Georgia, United States*, 15. https://doi.org/10.1109/WSC.2015.7408152

Geertsema, G., & Van den Brink, H. (2014, December). *Windkaart van nederland op 100 meter hoogte*. https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubTR/TR351.pdf

Gou, C.-X., Cai, B.-G., & Miao, Y. (2015). Multi-resolution model consistency maintenance method based on ontology mapping, 103–106. https://doi.org/10.1109/ICAIOT.2015.7111548

He, X., Chen, G., & Jin, G. (2014). Research on aggregation and disaggregation method of armored equipment motion model based on AHP. *Applied Mechanics and Materials*, *543-547*, 1734–1738. https://doi.org/10.4028/www.scientific.net/AMM.543-547.1734

IBM. (2021, March 4). *Time aggregation*. Retrieved April 21, 2022, from https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/tnpm/1.4.5?topic=series-time-aggregation

Jie, F., Li, F., Lu, Z., & Geng, B. (2012, January 1). Study of CGF task simulation system based on HLA. https://doi.org/10.1007/978-3-642-34396-4_44

Li, M., Bi, C., Liu, X., & Niu, D. (2008). Research on rules of aggregation and disaggregation for aircraft entity models in air force combat simulation. *2008 Asia Simulation Conference - 7th International Conference on System Simulation and Scientific Computing*, 1286–1290. https://doi.org/10.1109/ASC-ICSC.2008.4675569

Mao, H., Peng, G., & Zhang, H. (2013). Choosing models of appropriate resolutions for simulation: A MRM approach. *AsiaSim 2013* (pp. 188–199). Springer Berlin Heidelberg. Retrieved April 26, 2022, from http://link.springer.com/10.1007/978-3-642-45037-2_18

Morita, K. (2017). Reversible logic gates. In K. Morita (Ed.), *Theory of reversible computing* (pp. 77–101). Springer Japan. https://doi.org/10.1007/978-4-431-56606-9_4

Multi-model.nl. (2021). *Towards an open multi-model infrastructure*. https://multi-model.nl/

Nance, R. E. (1981). The time and state relationships in simulation modeling. *24*(4), 173–179. https://doi.org/10.1145/358598.358601

Natrajan, A., Jr, P., & Srinivasan, S. (1997). MRE: A flexible approach to multi-resolution modeling., 156–163. https://doi.org/10.1109/PADS.1997.594601

Nikolic, I., Warnier, M., Kwakkel, J. H., Chappin, E. J. L., Lukszo, Z., Brazier, F. M., Verbraeck, A., Cvetkovic, M., & Palensky, P. (2019). Principles, challenges and guidelines for a multi-model ecology. https://doi.org/10.4233/uuid:1aa3d16c-2acd-40ce-b6b8-0712fd947840

Overstreet, & Nance. (1986). World view based discrete event model simplification [ISBN: 9780444701305], *165-179*. https://simulation.tudelft.nl/dsol/manual/simulation-theory/introduction

Pandas development. (2022). *Pandas.DataFrame.rolling — pandas 1.4.3 documentation*. https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rolling.html

Perumalla, K. (2006). Parallel and distributed simulation: Traditional techniques and recent advances, 84–95. https://doi.org/10.1109/WSC.2006.323041

Quintel. (2021). *Energy transition model*. https://quintel.com/etm

Quintel. (2022a). *Energiegraaf/stromendiagram | ETM*. https://engine.energytransitionmodel.com/inspect/915321/layouts/energy

Quintel. (2022b). *Energy calculations | energy transition model*. https://docs.energytransitionmodel.com/main/energy-calculations

Quintel. (2022c, January 10). *ETSource github*. https://github.com/quintel/etsource

Rabelo, L., Kim, K., Park, T., Pastrana, J., Marin, M., Lee, G., Nagadi, K., Ibrahim, B., & Gutierrez, E. (2016). Multi resolution modeling, 2523–2534. https://doi.org/10.1109/WSC.2015.7408362

RES. (2021). *Nationaal Programma Regionale Energiestrategie*. https://www.regionale-energiestrategie.nl/default.aspx

Reynolds, P. F., & Natrajan, A. (1997). Consistency maintenance in multiresolution simulations. *ACM Transactions on Modeling and Computer Simulation*, *7*(3), 25. https://doi.org/10.1145/259207.259235

Reynolds, P. F., Natrajan, A., & Srinivasan, S. (1997). Consistency maintenance in multiresolution simulation. *ACM Transactions on Modeling and Computer Simulation*, *7*(3), 368–392. https://doi.org/10.1145/259207.259235

RVO. (2021, April 20). *Transitievisie Warmte en Wijkuitvoeringsplan*. https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/duurzame-energie-opwekken/aardgasvrij/transitievisie-warmte-en-wijkuitvoeringsplan

Salome, S. (2021). *On the challenge of designing a robust military force: A multi-resolution modelling approach to improve the performance of a naval force support system*. Delft, Netherlands. https://repository.tudelft.nl/islandora/object/uuid%3Abaa50bd3-e32a-44ee-9fe4-f9593d3e0829

Schulze, M. (2000). Linear programming for optimization. https://www.researchgate.net/publication/2420905_Linear_Programming_for_Optimization

Seck, M. D., & Honig, H. J. (2012). Multi-perspective modelling of complex phenomena. *Computational and Mathematical Organization Theory*, 128–144. https://doi.org/10.1007/s10588-012-9119-9

Skogan, D. (2001). Managing resolution in multi-resolution databases. *In Proceedings of the 8th Scandinavian Research Conference*, 99–113. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.2283&rep=rep1&type=pdf

Stell, J., & Worboys, M. (1998). Stratified map spaces: A formal basis for multi-resolution spatial databases. *Proceedings of the 8th International Symposium on Spatial Data Handling*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.50.2853&rep=rep1&type=pdf

Tan, G., Ng, W. N., & Moradi, F. (2001). Aggregation/disaggregation in HLA multi-resolution distributed simulation. https://doi.org/10.1109/DISTRA.2001.946446

TenneT. (2021a). *Balanceringsverantwoordelijkheid*. https://www.tennet.eu/nl/elektriciteitsmarkt/nederlandse-markt/balanceringsverantwoordelijkheid/

TenneT. (2021b, April). *Integrale infrastructuurverkenning 2030-2050*. https://www.tennet.eu/nl/bedrijf/publicaties/ii3050/

Tolk, A. (2003). *Levels of conceptual interoperability*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.196.1923&rep=rep1&type=pdf

U.S. Air Force. (2015, September). *F-22 raptor*. https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104506/f-22-raptor/

Vangheluwe, H., de Lara, J., & Mosterman, P. J. (2002). An introduction to multi-paradigm modelling and simulation. *Proceedings of the AIS'2002 Conference - Lisbon, Portugal*, 12. https://www.researchgate.net/publication/243776266_An_introduction_to_multi-paradigm_modelling_and_simulation

Xuefei, Y., Qiang, L., Xiaolong, W., Dong, L., & Shoubiao, W. (2017). Non-consistence aggregation-disaggregation technology for battle simulation study of SoS. *Proceedings of 2017 the 7th International Workshop on Computer Science and Engineering*. https://doi.org/10.18178/wcse.2017.06.066

Yilmaz, L., Lim, A., Bowen, S., & Ören, T. (2007). Requirements and design principles for multisimulation with multiresolution, multistage multimodels. https://doi.org/10.1109/WSC.2007.4419678

Zhang, X.-Y., Trame, M., Lesko, L., & Schmidt, S. (2015). Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models. *CPT: Pharmacometrics & Systems Pharmacology*, *4*(2), 69–79. https://doi.org/10.1002/psp4.6

Zhou, H., Chen, B., Han, Z. X., & Zhang, F. Q. (2009). Study on probability distribution of prices in electricity market: A case study of zhejiang province, china. *Communications in Nonlinear Science and Numerical Simulation*, *14*(5), 2255–2265. https://doi.org/10.1016/j.cnsns.2008.04.020

A

Appendix, case study 1

This appendix contains additional tables, graphs and clarification for case study 1 (ETM-HWP coupling. The HWP model, (dis)aggregation functions and sensitivity analyses used are available on Github. The ETM model is available online.

A.1. HWP sensitivity analysis, coarse data

Cashflow	Power +	Power o	Power -
Eprice+	1.21110	1.10105	0.99100
Epriceo	1.09994	1.00000	0.90005
Eprice-	0.98878	0.89895	0.80910

TMK	Power +	Power o	Power -
Eprice +	1.01056	1	0.98728
Eprice o	1.01056	1	0.98727
Eprice -	1.01056	1	0.98727

Figure A.1: HWP coarsened input sensitivity analysis

ΔCashflow	Power +	Power o	Power -
Eprice+	1.03482	1	0.95941
Epriceo	1.03482	1	0.95941
Eprice-	1.03482	1	0.95941

Figure A.2: Windfarm cashflow with and without a hydrogen buffer (coarsened data)

A.2. Case 1 individual wind speed disaggregation technique results

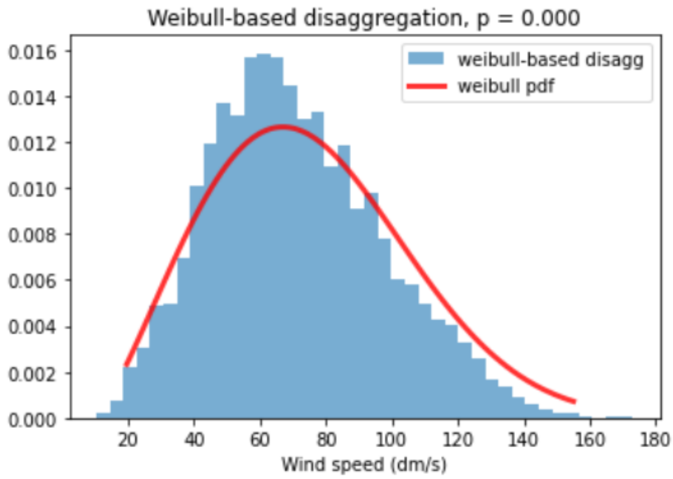

Figure A.3: Weibull-based disaggregation with accompanying kolmogorov-smirnov test p-value

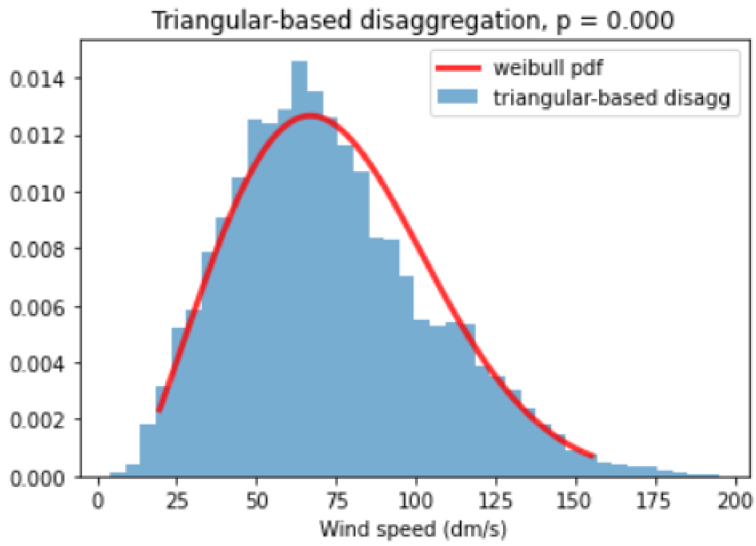

Figure A.4: Triangular-based disaggregation with accompanying kolmogorov-smirnov test p-value

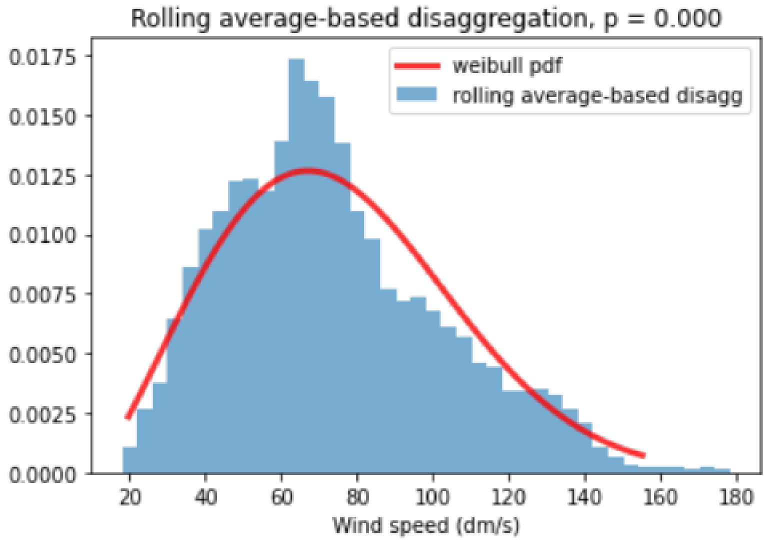

Figure A.5: Rolling average-based disaggregation with accompanying kolmogorov-smirnov test p-value

A.3. Case 1 data samples from each disaggregation technique

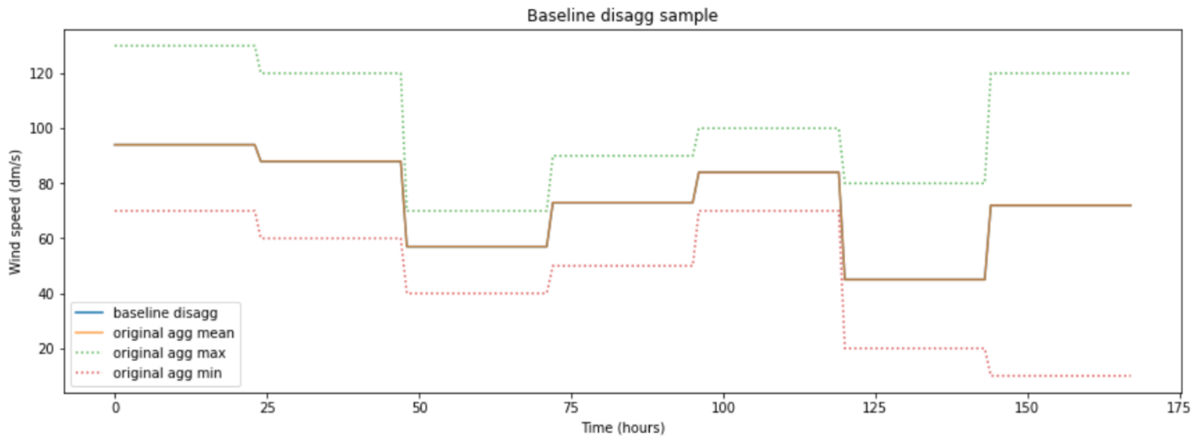

Figure A.6: One week data sample of baseline disaggregation

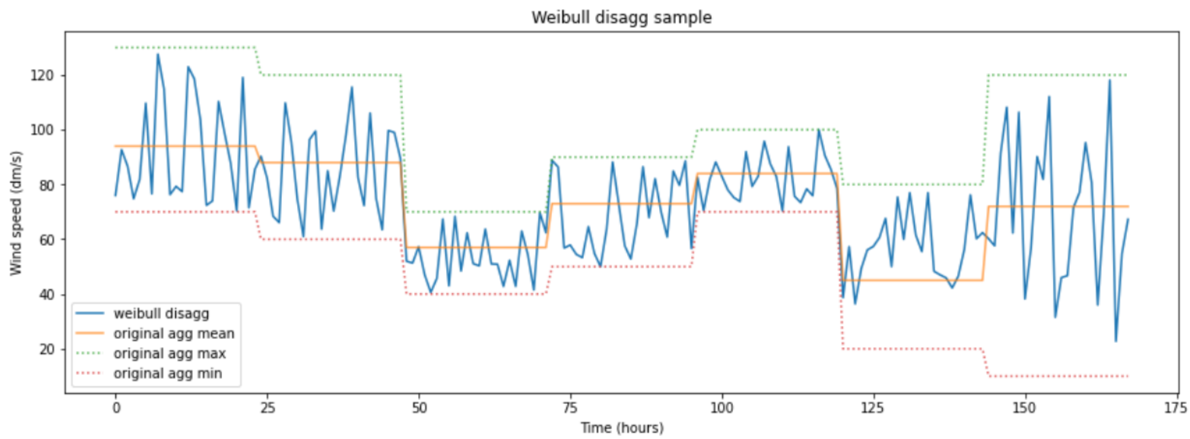

Figure A.7: One week data sample of weibull disaggregation

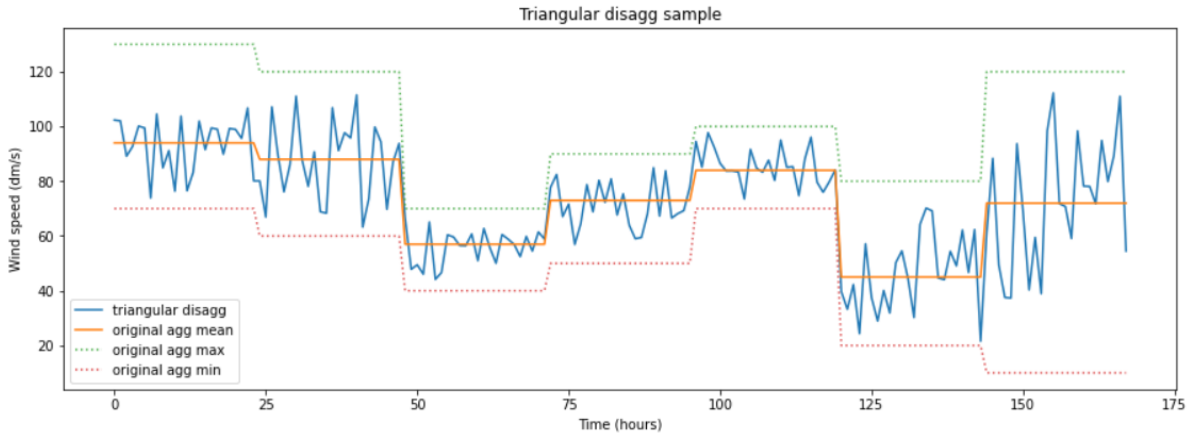

Figure A.8: One week data sample of triangular disaggregation

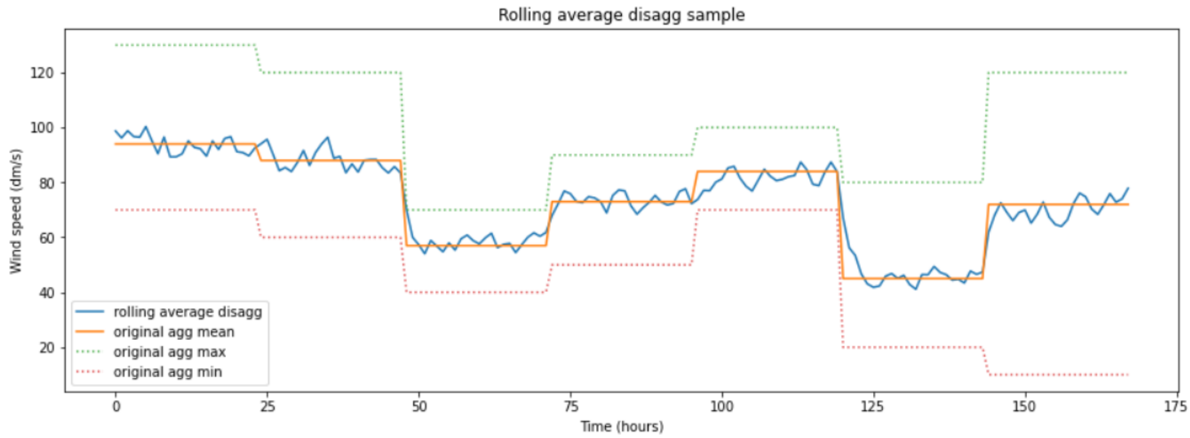

Figure A.9: One week data sample of rolling average disaggregation

A.4. Case 1 thrashing test results for each disaggregation technique

Figures A.12, A.13 and A.14 show no meaningful deviation from the original value and can thus be regarded as resistant to thrashing. One interesting result will be analysed further here, namely the results of the triangular disaggregation method shown in figure A.10.

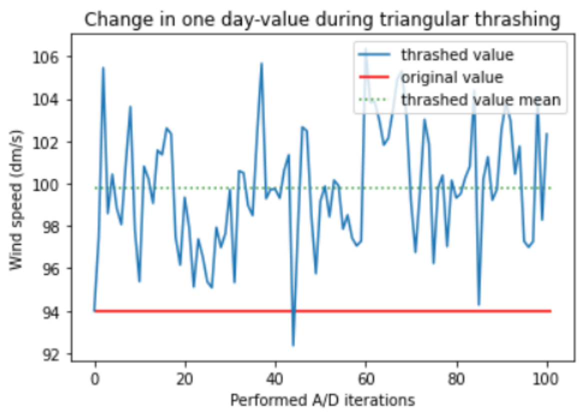

Figure A.10: ETM HWP thrashing test, triangular disaggregation method

Triangular distributions are widely used, especially in situations where limited data exists to base distributions on. Unfortunately, figure A.10 shows that out of the tested distributions shown in appendix A.4, the popular triangular is the only one to show a structural drift away from the mean after repeated A/D.

The triangular distribution seems to be by its nature inherently vulnerable to a structural bias during thrashing. During thrashing a new disaggregate is created from which a new aggregate will again be made, continuing the cycle for a given amount of iterations. Given an e.g. right skewed triangular distribution (where the max is further from the mean than the min), the values resulting from disaggregation will lean towards the max, thus moving the mean to the max as well for the next disaggregation. This process is expected to continue until the probability density (pdf) function resembles an isosceles triangle. Once this pdf shape is reached, the values are expected to stabilize. Figure A.10 shows this is the case. This particular distribution is skewed towards the max (min: 70, mean: 94, max: 130), thus moving the mean of aggregates gradually until an isosceles triangle pfd shape is reached around 100 dm/s. Figure A.11 adds to this issue by showing that tightening the consistency threshold further does not fix the problem. It merely increases the amount of iterations needed to reach the isosceles shape, thus slowing the drift away from the original value.

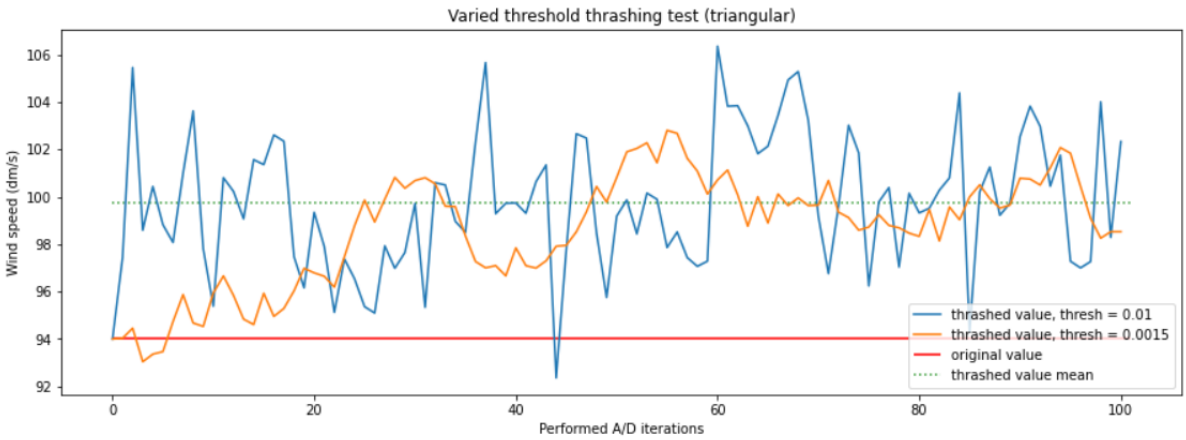

Figure A.11: Triangular varied threshold thrashing test: ±1%, ±0.15%

With all this being said, by no means does a thrashing issue outright disqualify the triangular distribution from use in A/D efforts. Multi-models that communicate infrequently will suffer less from the thrashing-induced drift. Given that the hypothetical case 1 coupling as defined in section 3.4.1 only communicates *once* at the start of the run, the negative effects of thrashing will remain limited. Do note that the drift is still present, even if a disaggregation only occurs once.

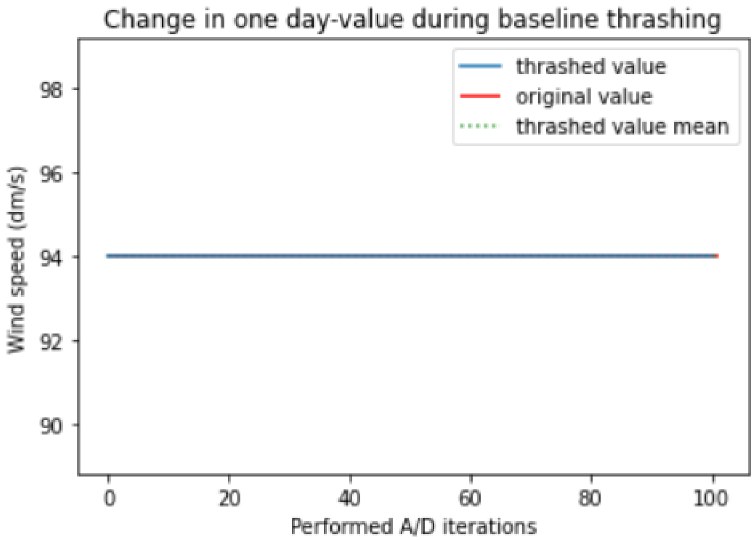

Figure A.12: Baseline method thrashing test through repeated disagg (N = 100)

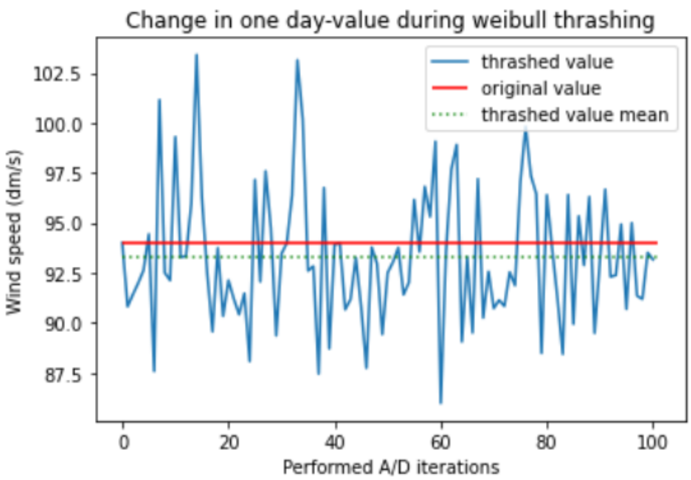

Figure A.13: Weibull method thrashing test through repeated disagg (N = 100)

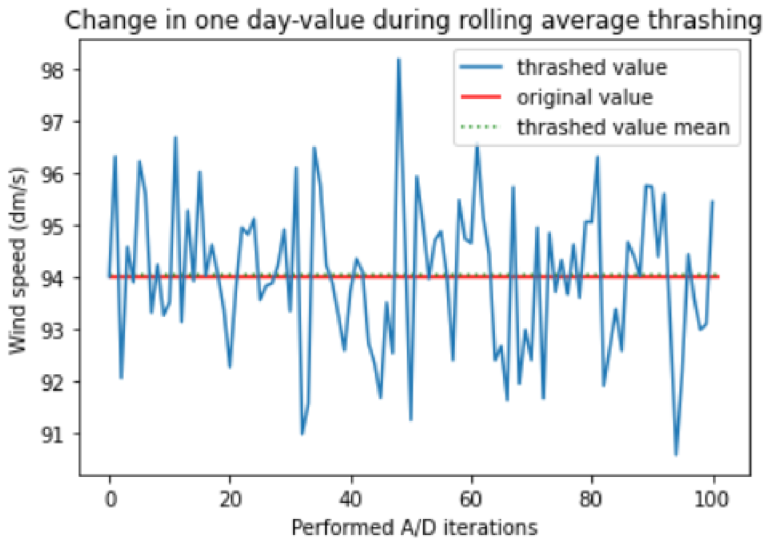

Figure A.14: Rolling average method thrashing test through repeated disagg (N = 100)

A.5. Case 1 sensitivity analysis results for each disaggregation technique

Cashflow	Power +	Power o	Power -
Eprice+	1.2117	1.1025	0.9932
Epriceo	1.0990	1.0000	0.9009
Eprice-	0.9863	0.8975	0.8087

TMK	Power +	Power o	Power -
Eprice +	1.0024	1.0000	0.9972
Eprice o	1.0024	1.0000	0.9972
Eprice -	1.0024	1.0000	0.9972

Figure A.15: Case 1 input sensitivity analysis, baseline disagg

Cashflow	Power +	Power o	Power -
Eprice+	1.2120	1.1028	0.9936
Epriceo	1.0989	1.0000	0.9010
Eprice-	0.9858	0.8972	0.8085

TMK	Power +	Power o	Power -
Eprice +	1.0018	1.0000	0.9981
Eprice o	1.0018	1.0000	0.9981
Eprice -	1.0018	1.0000	0.9981

Figure A.16: Case 1 input sensitivity analysis, weibull disagg

Cashflow	Power +	Power o	Power -
Eprice+	1.2117	1.1025	0.9931
Epriceo	1.0990	1.0000	0.9009
Eprice-	0.9863	0.8975	0.8087

TMK	Power +	Power o	Power -
Eprice +	1.0012	1.0000	0.9982
Eprice o	1.0012	1.0000	0.9982
Eprice -	1.0012	1.0000	0.9982

Figure A.17: Case 1 input sensitivity analysis, triangular disagg

Cashflow	Power +	Power o	Power -
Eprice+	1.2117	1.1025	0.9932
Epriceo	1.0990	1.0000	0.9009
Eprice-	0.9863	0.8975	0.8087

TMK	Power +	Power o	Power -
Eprice +	1.0019	1.0000	0.9987
Eprice o	1.0019	1.0000	0.9987
Eprice -	1.0019	1.0000	0.9987

Figure A.18: Case 1 input sensitivity analysis, rolling average disagg

B

Appendix, case study 2

This appendix contains additional tables, graphs and clarification for case study 2 (ETM-EVM coupling. The EVM model, (dis)aggregation functions and sensitivity analyses used are available on Github. The ETM model is available online.

B.1. Case 2 time aggregation consistency function implementation

As documented by Zhou et al. (2009), a normal distribution is set as the standard for predicting electricity prices. However, when examining the electricity prices from ETM in figure B.1, it does not seem to fit well to a normal distribution. Rather the ETM has certain discrete price levels that it steps between, as seen in appendix B.5. The ETM model audit (4.1) suggests that this could be a result of the ETM's aggregation of sectors into a single node on a graph, such as types of electricity producers. This then in a more homogeneous marginal cost curve of electricity.

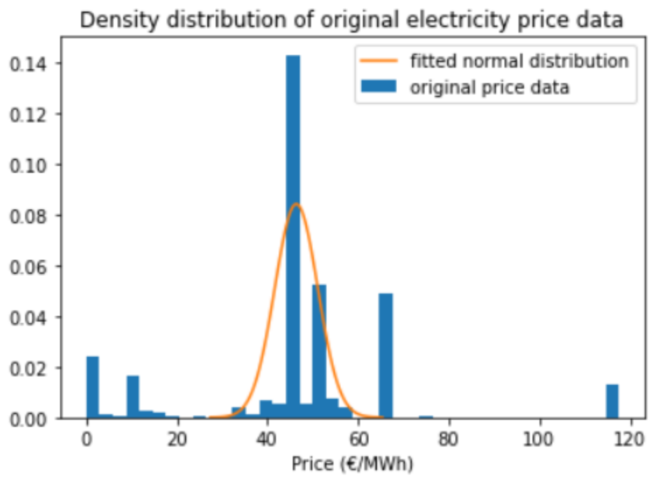

Figure B.1: case 2, normal distribution fit to price data

Discrete data such as seen in figure B.1 is difficult to match to a (normal) distribution. However, it might not be necessary to mimic the original data in the conventional sense (as described in equation 4.8). Section 4.3.2 of the EVM model audit states that EV-agents use the time series to decide when is the cheapest time to charge. Crucially, the EV-agents then only use the relative difference in prices to base behaviour on, not the prices themselves. Therefore the direction of the derivative of the prices is the important facet to translate during disaggregation. Equations B.1 and B.2 provide a concrete implementation of this concept. Instead of abiding by and equal average rule, the disaggregation method can be tested on how well it translates the differences in price from hour to hour. In order to check if the disaggregation abides by the threshold, equation B.2 translated the aggregated and disaggregated datasets each to a list of -1, 0, 1 values to indicate a negative, flat or positive change (the size does not matter to the EV-agents). Then equation B.1 provides a similarity factor by summing up the number of times that the direction of the aggregate and disaggregate are the same.

$$\frac{\int_{i=1}^{N} Diff_{price\ disagg_i} = Diff_{price\ agg_i}}{N} \geq E_{thresh} \tag{B.1}$$

$$Diff = \begin{cases} -1\ if\ (average)\ price\ hour_i & < (average)\ price\ hour_{i-1} \\ 0\ if\ (average)\ price\ hour_i & = (average)\ price\ hour_{i-1} \\ 1\ if\ (average)\ price\ hour_i & > (average)\ price\ hour_{i-1} \end{cases} \tag{B.2}$$

B.2. Case 2 electricity price reverse rolling average function accuracy test

Given a centered rolling average method (Pandas development, 2022), a list of numbers $X_1 - X_N$ and its rolled average $average_{X_{1,N}}$, the last value X_N in the list can be calculated using equation B.3. Once the calculation of X_N is done, the equation can be used on the list $X_2 - X_N + 1$ etc to reverse the rolling average process. However, equation B.3 requires the specific inputs of $X_1 - X_{N-1}$ to complete the first re-aggregation. For the centered rolling average technique, the list of first numbers that would have to be stored equals half the window size. It is not uncommon to have to store some information to perform a reversion calculation. The need to store a small set of 'junk' information to perform the aggregation bares resemblance to reversible computing, where 'junk bits' are often a byproduct of rudimentary reversible logic gates (Morita, 2017).

$$X_n = (N * average_{X_{1,N}}) - \sum_{i=1}^{N-1} X_i \tag{B.3}$$

Equation B.3 is used to great effect in thrashing tests for the case 2 price data. Figure B.2 shows that values remain near constant through disaggregation. This is the case for all window sizes (accompanying graphs available in appendix B.3). Therefore the current price disaggregation technique can be viewed as very well resistant to thrashing. All spatial disaggregation techniques also show no structural drift during thrashing. These methods can therefore also be viewed as resistant to thrashing (see appendix B.4 for all figures).

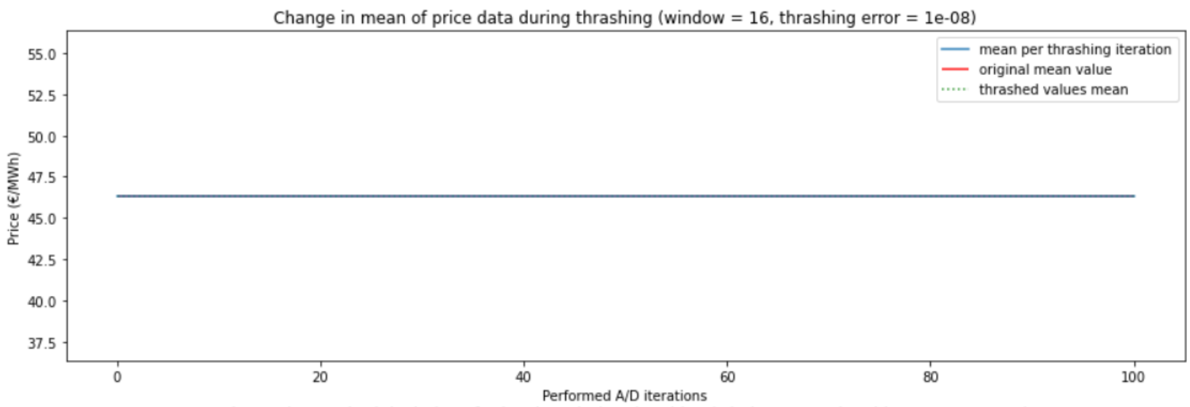

Figure B.2: Electricity price rolling average thrashing test through repeated disagg (N = 100, window = 16)

As stated above, the thrashing function used in these tests rely on a reverse rolling-average function. Given the finite accuracy of the python environment which the function was built in, possible errors in the reversal function could enlarge as the reversal goes. Two tests are performed to see if there is a significant problem. Figure B.3 shows the progression of the mean error by comparing a re-aggregated price data set to the original price data set. The test in figure B.4 performs the same test on data that was first thrashed 100 times. As expected, the error grows (and does not decrease) the further the function goes on. Fortunately, both tests show that the generated error is small enough relative to the price values (< 0.001%) to be deemed insignificant. If future models use datasets that are significantly larger, the error might eventually become large enough to be significant. That would reduce the ability of the reversal function during thrashing test use. Although, given the current error size, the error is not expected to become significant even for datasets that are some orders of magnitude larger.

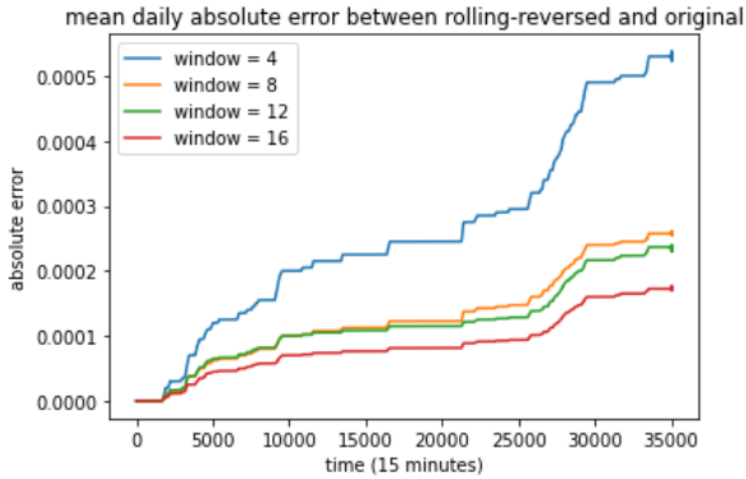

Figure B.3: Reverse rolling average function error development

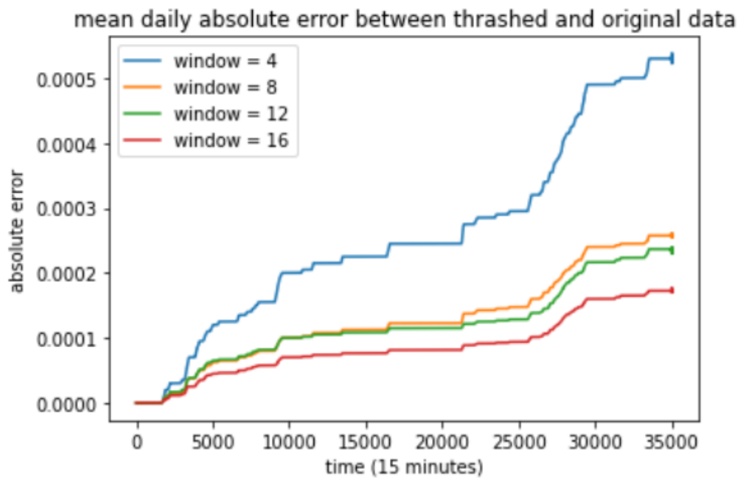

Figure B.4: Reverse rolling average function error development, after 100 thrashing iterations

B.3. Case 2 electricity price thrashing test of rolling average disaggregation at various windows

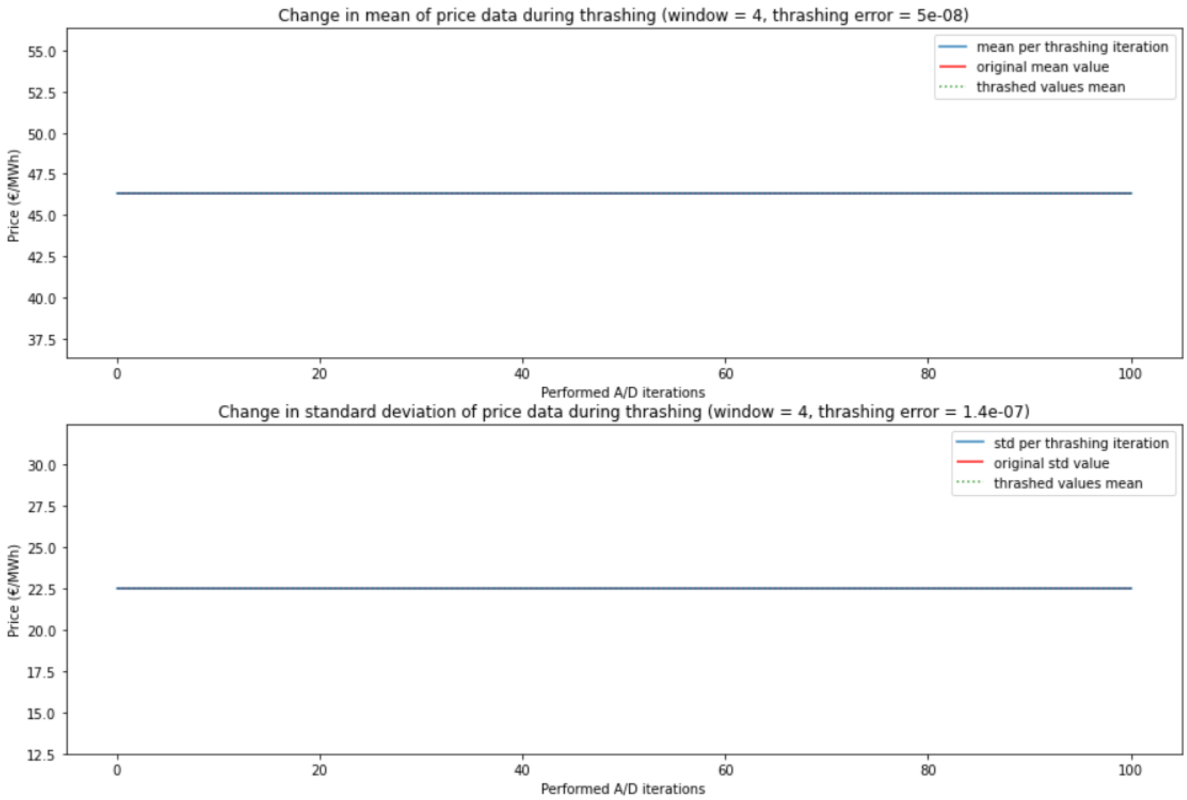

Figure B.5: Electricity price rolling average thrashing test through repeated disagg (N = 100, window = 4)

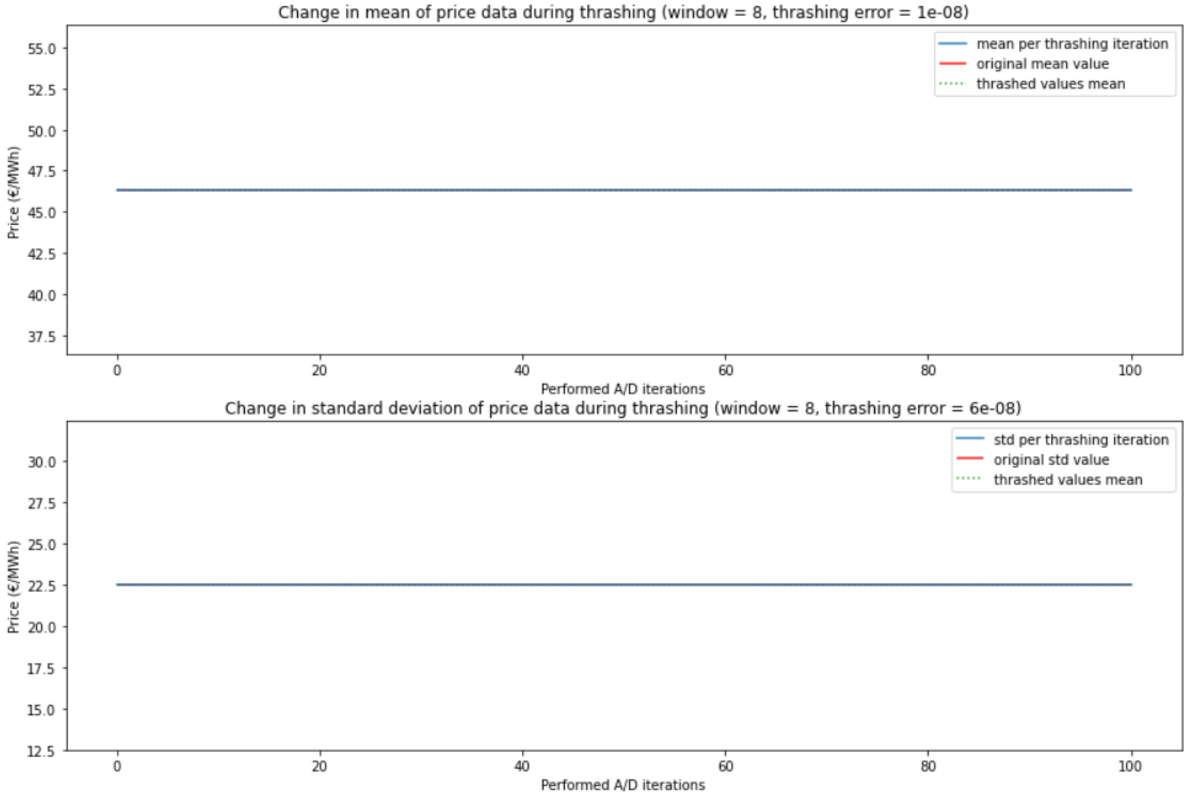

Figure B.6: Electricity price rolling average thrashing test through repeated disagg (N = 100, window = 8)

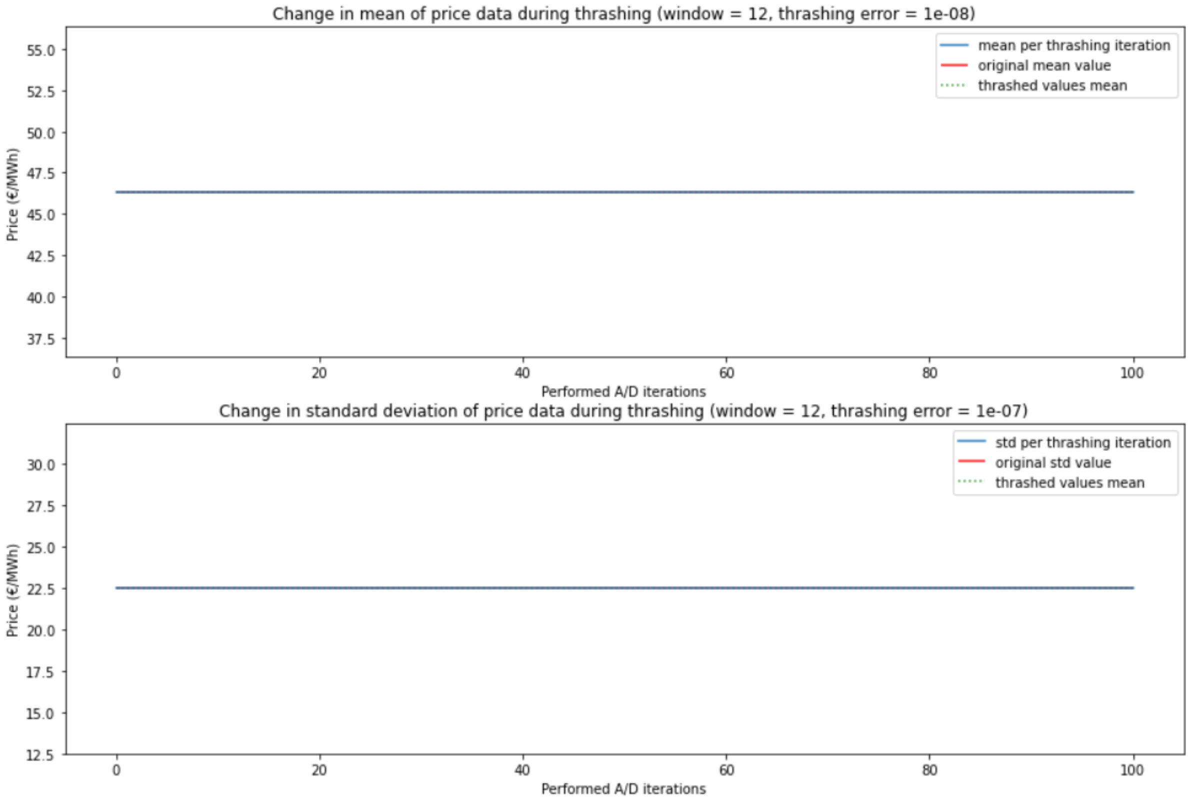

Figure B.7: Electricity price rolling average thrashing test through repeated disagg (N = 100, window = 12)

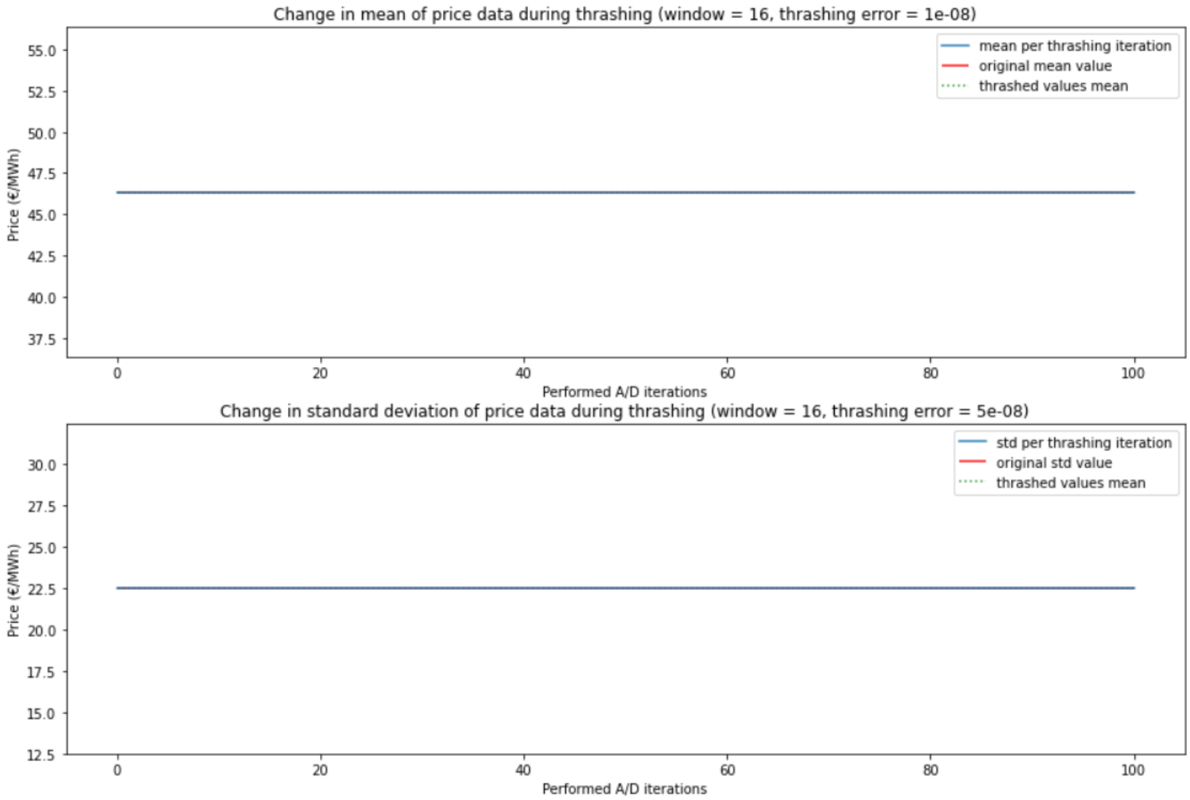

Figure B.8: Electricity price rolling average thrashing test through repeated disagg (N = 100, window = 16)

B.4. Case 2 spatial thrashing tests

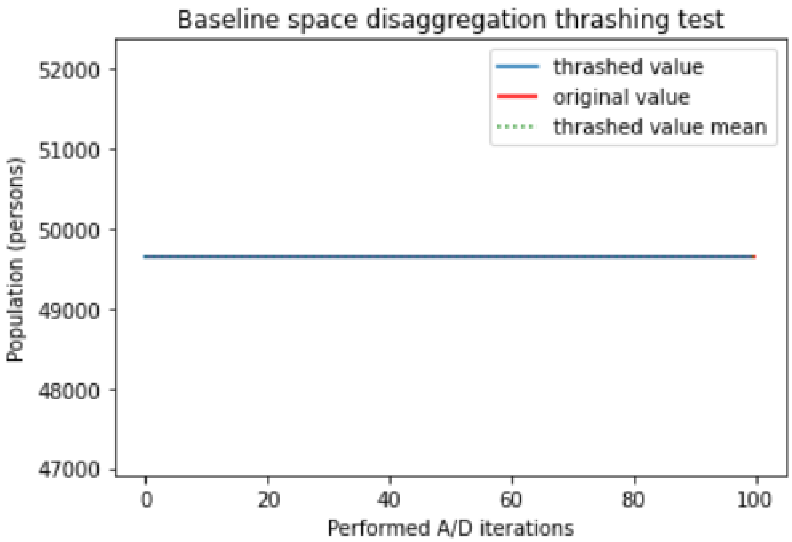

Figure B.9: Change in one region population count during baseline method thrashing

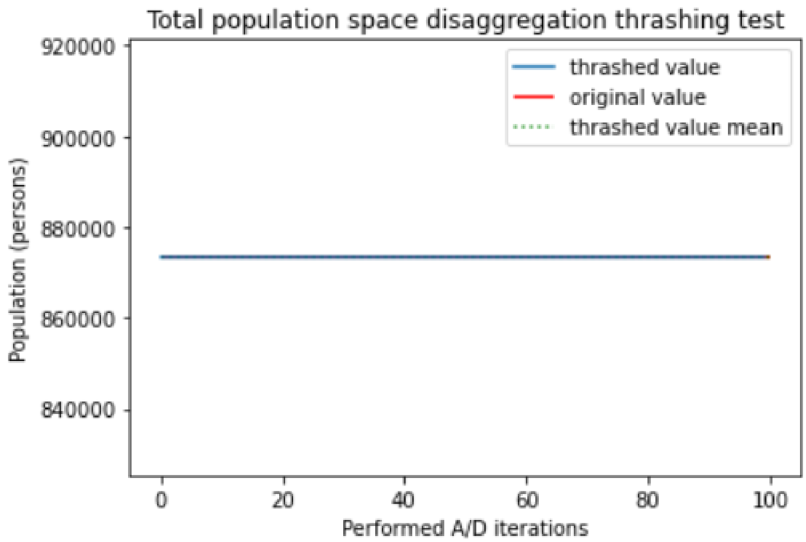

Figure B.10: Change in one region population count during total population-method thrashing

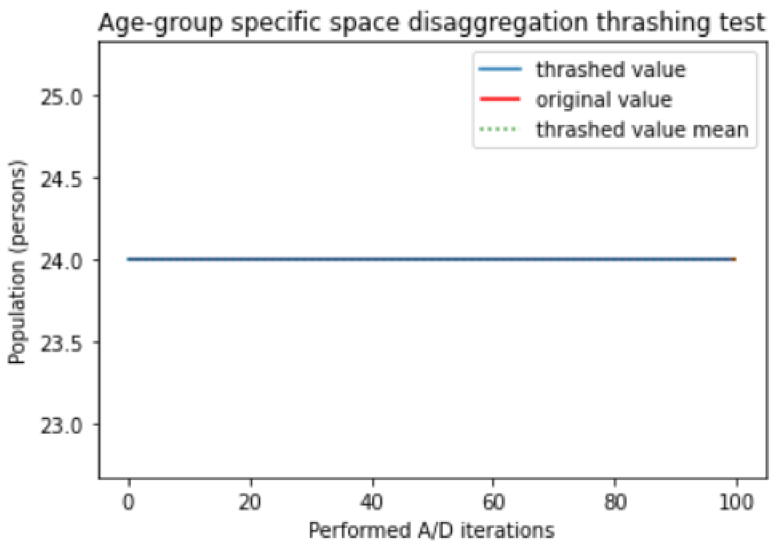

Figure B.11: Change in one region population count during age group-specific method thrashing

B.5. Case 2 electricity price samples

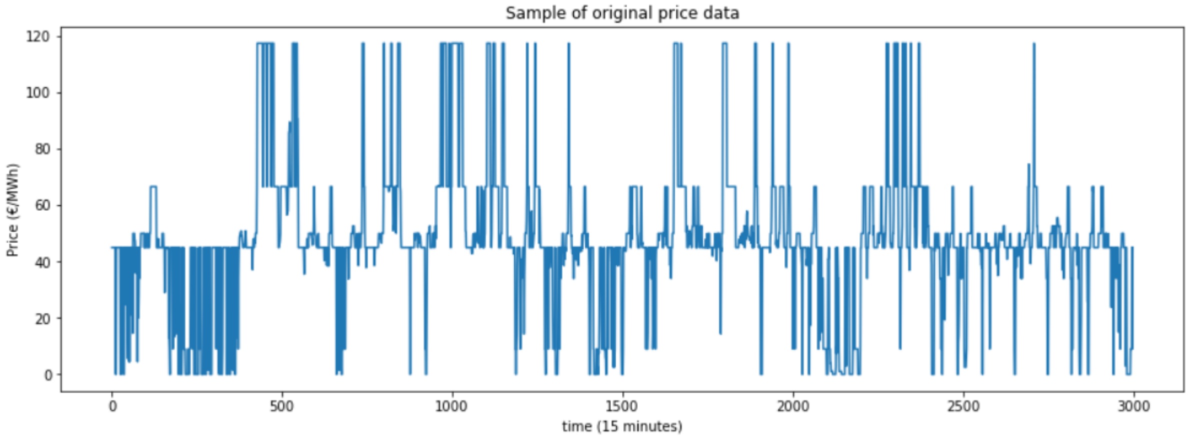

Figure B.12: Sample of original price data, ETM

B.6. Case 2 spatial population distribution visualisations

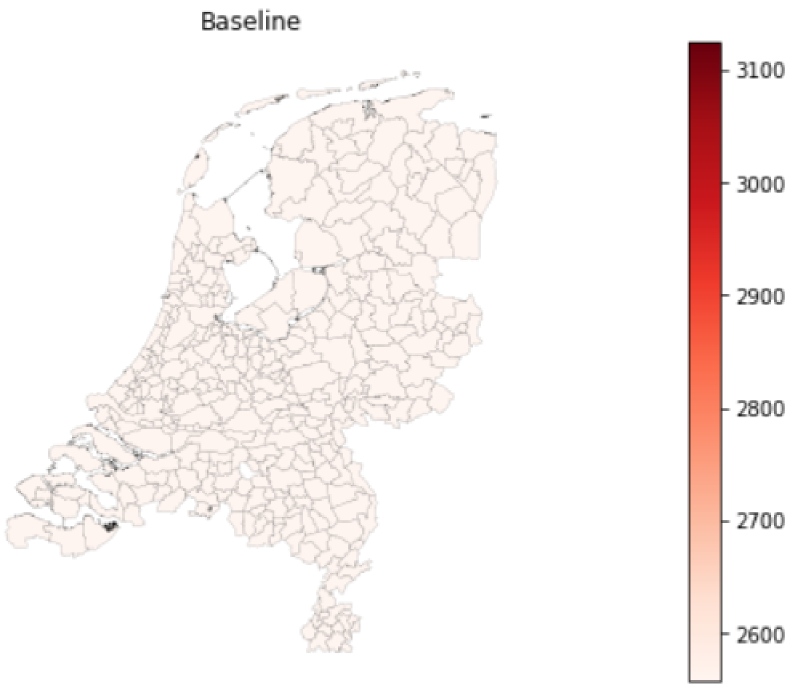

Figure B.13: Baseline spatial population distribution

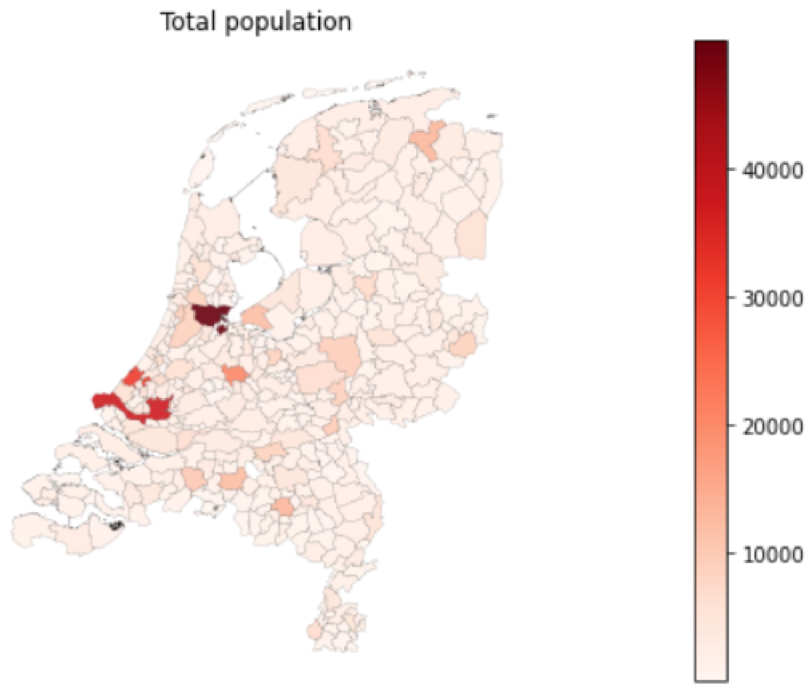

Figure B.14: Total population spatial population distribution

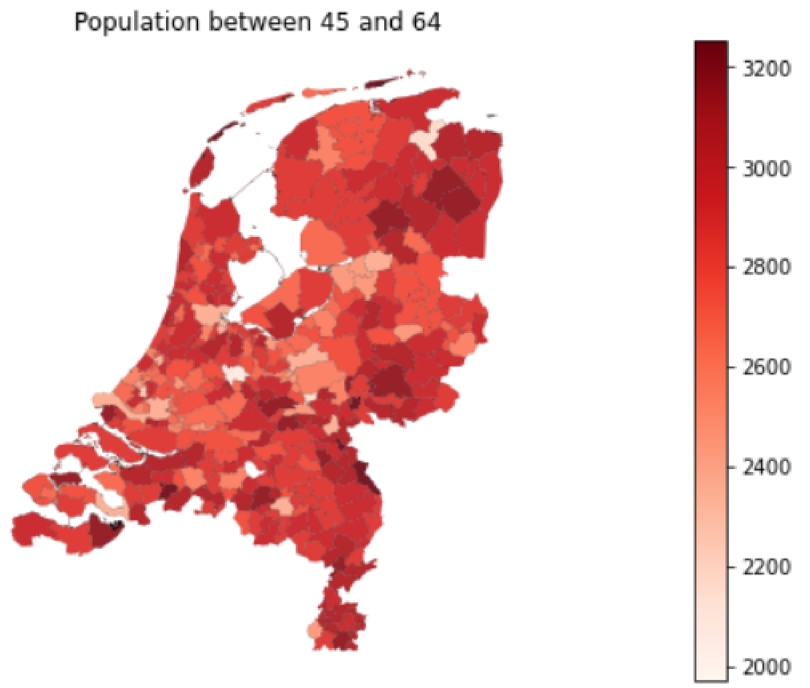

Figure B.15: Age-group specific spatial population distribution

B.7. Case 2 sensitivity analysis results for each disaggregation technique

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4539.9	4130.9	3713
Epriceo	4539.9	4130.9	3713
Eprice-	4539.9	4130.9	3713

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.899
Epriceo	1.099	1.000	0.899
Eprice-	1.099	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	67839.14	61955.95	55855.48
Epriceo	67839.14	61955.95	55855.48
Eprice-	67839.14	61955.95	55855.48

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.095	1.000	0.902
Epriceo	1.095	1.000	0.902
Eprice-	1.095	1.000	0.902

Figure B.16: Case 2 input sensitivity analysis, price baseline disagg

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4520.6	4111.5	3694.8
Epriceo	4520.6	4111.5	3694.8
Eprice-	4520.6	4111.5	3694.8

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.100	1.000	0.899
Epriceo	1.100	1.000	0.899
Eprice-	1.100	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	59911.36	54772.51	49266.2
Epriceo	59911.36	54772.51	49266.2
Eprice-	59911.36	54772.51	49266.2

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.899
Epriceo	1.094	1.000	0.899
Eprice-	1.094	1.000	0.899

Figure B.17: Case 2 input sensitivity analysis, price rolling average disagg (window = 4)

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4519.5	4111.02	3694.7
Epriceo	4519.5	4111.02	3694.7
Eprice-	4519.5	4111.02	3694.7

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.899
Epriceo	1.099	1.000	0.899
Eprice-	1.099	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	59754.21	54640.4	49145.17
Epriceo	59754.21	54640.4	49145.17
Eprice-	59754.21	54640.4	49145.17

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.899
Epriceo	1.094	1.000	0.899
Eprice-	1.094	1.000	0.899

Figure B.18: Case 2 input sensitivity analysis, price rolling average disagg (window = 8)

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4519.5	4111	3694.7
Epriceo	4519.5	4111	3694.7
Eprice-	4519.5	4111	3694.7

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.899
Epriceo	1.099	1.000	0.899
Eprice-	1.099	1.000	0.899

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	59754.21	54640.4	49145.17
Epriceo	59754.21	54640.4	49145.17
Eprice-	59754.21	54640.4	49145.17

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.899
Epriceo	1.094	1.000	0.899
Eprice-	1.094	1.000	0.899

Figure B.19: Case 2 input sensitivity analysis, price rolling average disagg (window = 12)

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4510.4	4105	3688.3
Epriceo	4510.4	4105	3688.3
Eprice-	4510.4	4105	3688.3

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.898
Epriceo	1.099	1.000	0.898
Eprice-	1.099	1.000	0.898

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	60292.9	55116.07	49588.1
Epriceo	60292.9	55116.07	49588.1
Eprice-	60292.9	55116.07	49588.1

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.094	1.000	0.900
Epriceo	1.094	1.000	0.900
Eprice-	1.094	1.000	0.900

Figure B.20: Case 2 input sensitivity analysis, price rolling average disagg (window = 16)

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4642.4	4248.7	3840.3
Epriceo	4642.4	4248.7	3840.3
Eprice-	4642.4	4248.7	3840.3

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.093	1.000	0.904
Epriceo	1.093	1.000	0.904
Eprice-	1.093	1.000	0.904

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	59723.02	55185.1	50095.65
Epriceo	59723.02	55185.1	50095.65
Eprice-	59723.02	55185.1	50095.65

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.082	1.000	0.908
Epriceo	1.082	1.000	0.908
Eprice-	1.082	1.000	0.908

Figure B.21: Case 2 input sensitivity analysis, space baseline disagg

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4523.7	4101.3	3696.7
Epriceo	4523.7	4101.3	3696.7
Eprice-	4523.7	4101.3	3696.7

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.103	1.000	0.901
Epriceo	1.103	1.000	0.901
Eprice-	1.103	1.000	0.901

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	59234.74	53915.2	49077.05
Epriceo	59234.74	53915.2	49077.05
Eprice-	59234.74	53915.2	49077.05

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.099	1.000	0.910
Epriceo	1.099	1.000	0.910
Eprice-	1.099	1.000	0.910

Figure B.22: Case 2 input sensitivity analysis, space total population disagg

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	4638	4266.9	3818.9
Epriceo	4638	4266.9	3818.9
Eprice-	4638	4266.9	3818.9

Mean national power demand	Evs +	Evs o	Evs -
Eprice+	1.087	1.000	0.895
Epriceo	1.087	1.000	0.895
Eprice-	1.087	1.000	0.895

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	60538.81	55206.55	50074.32
Epriceo	60538.81	55206.55	50074.32
Eprice-	60538.81	55206.55	50074.32

Mean national VTG capacity	Evs +	Evs o	Evs -
Eprice+	1.097	1.000	0.907
Epriceo	1.097	1.000	0.907
Eprice-	1.097	1.000	0.907

Figure B.23: Case 2 input sensitivity analysis, space age specific disagg

B.8. Case 2 sensitivity analysis results for one municipality

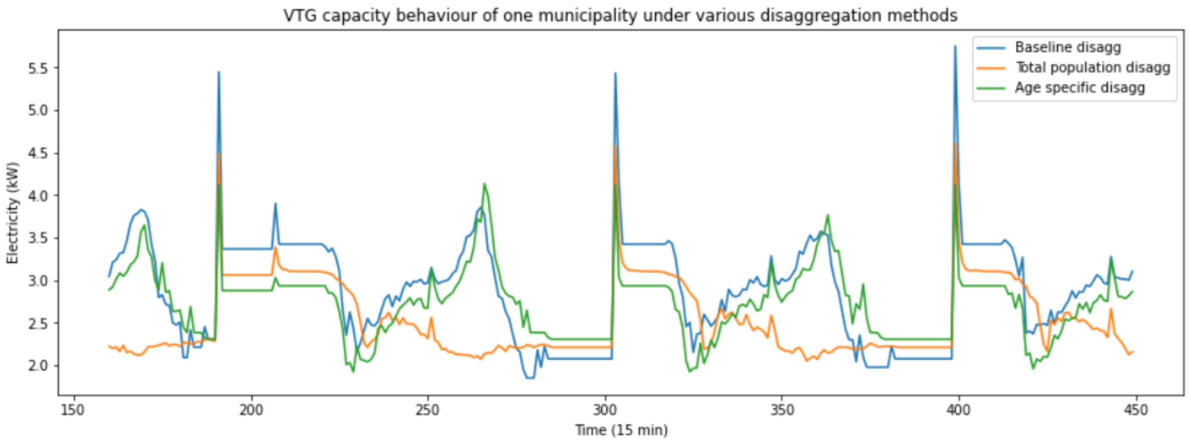

Figure B.24: VTG capacity of municipality 'Amsterdam'

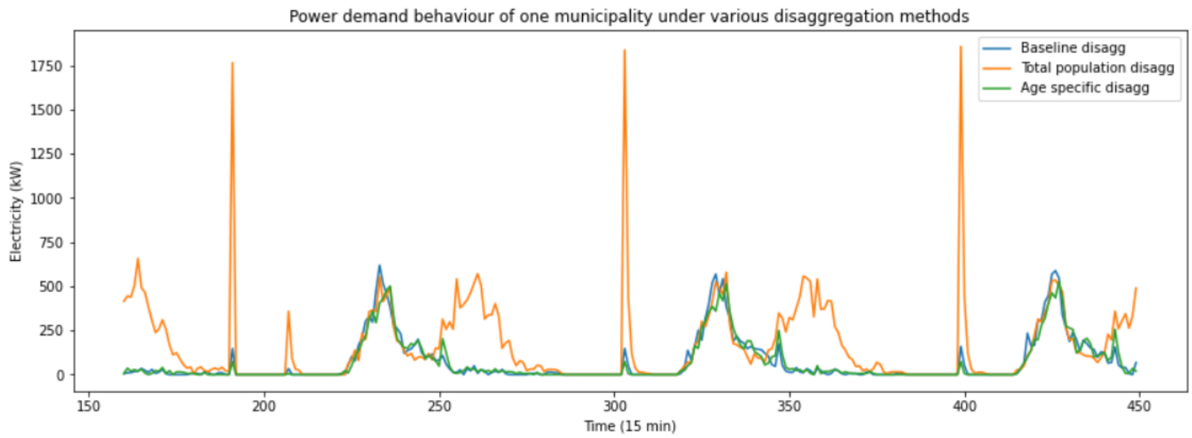

Figure B.25: Power demand of municipality 'Amsterdam'

B.9. Case 2 samples of final multi-model run

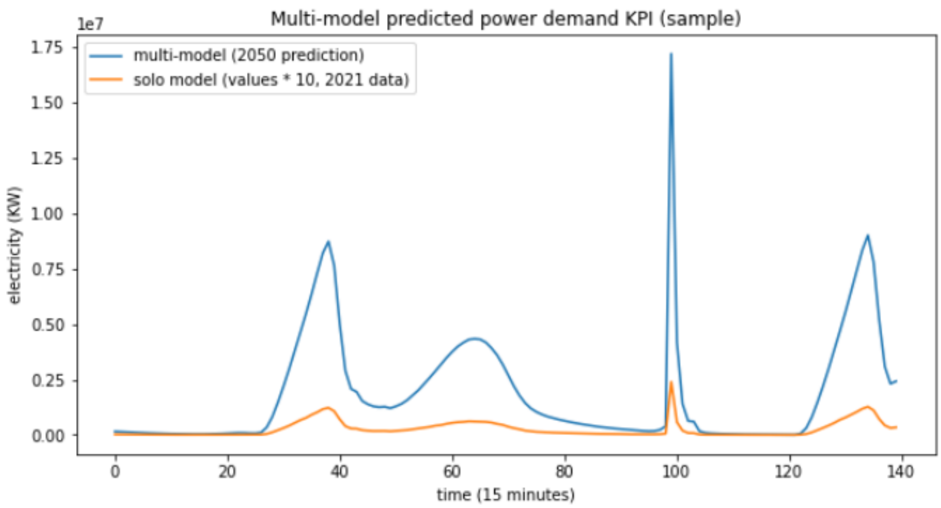

Figure B.26: Sample of total power KPI during multi-model run

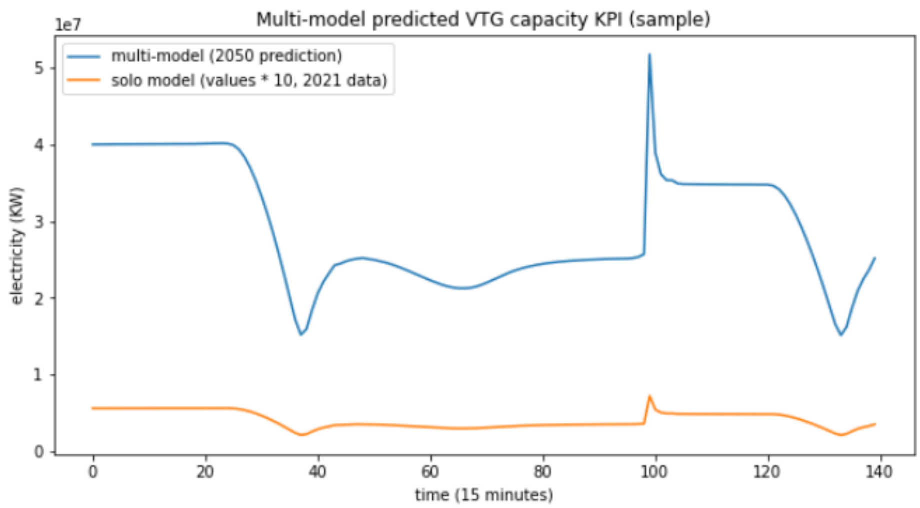

Figure B.27: Sample of VTG KPI during multi-model run