Exploring the Potential of Down-Sampling to Reduce Temporal Resolution in Energy System Optimisation Models

Master Thesis Energy Science

Laura C. Brafine

Student ID: 6003672
l.c.brafine@students.uu.nl

Supervisors:

Dr. Matteo Gazzani[1]
Msc. Zhi Gao[1]
Dr. Germán Morales-España[2,3]

[1]University of Utrecht, Utrecht, The Netherlands
[2]TNO, Amsterdam, The Netherlands
[3]Delft University of Technology, Delft, The Netherlands

September 13, 2024

Abstract

Energy system optimisation models (ESO) are increasingly important for managing renewable energy source (RES) integration. However, due to the intermittency of RES, the inclusion of storage and the increasing interconnectedness of energy sectors the complexity of these models increases and makes them computationally intensive. This study aims to assess the effectiveness of alternative down-sampling (DS) methods in reducing temporal resolution in ESO models while approximating accuracy on hourly resolution. We compare this performance to benchmark DS that uses the default average or energy value for aggregation. Basic statistics and hybrid DS methods were applied in an operational model and two variations of an investment model within the context of a Northwestern European case study. The methods were evaluated based on their ability to reduce solve time and approximate model accuracy in terms of total system costs, investment decisions, flows and net position behaviour. The results show that all DS methods significantly reduced solve time across all model configurations, with reductions ranging from 44-75 % on 2-hourly resolution, 62-88 % on 3-hourly resolution and 81-92 % on 4-hourly resolution. Using the first, last, or midpoint values when aggregating, consistently outperformed the benchmark showing lower differences with the hourly reference for total system costs and investment decisions. This performance can be linked to their ability to represent energy value and ramp distributions at a lower resolution, which showed to be consistently better than the benchmark. Whereas using the maximum and minimum methods when aggregating, showed higher errors in energy value and ramp distribution representation and accordingly lower model performance in the evaluated metrics than the benchmark. The hybrid method attempts to balance energy value and ramp distribution representation by combining the benchmark average and maximum when aggregating. The hybrid method generally showed promising results in the different model configurations when applied to the energy demand profile and wind profiles, especially in total system costs and investment decisions. These findings suggest that alternative DS methods have the potential to offer a more effective solution than the benchmark that uses default averaging to reduce model complexity while approximating accuracy on hourly resolution. Moreover, they contribute to understanding the impact of DS on model performance. Future work could explore their application in larger-scale models, across other optimisation-based domains, in existing methods that use DS or applications that rely on time series such as time series forecasting, trend and anomaly detection.

Keywords:
Energy System Optimisation, Time Series Aggregation, Computational Complexity Reduction, Statistical Down-Sampling, Adaptive Down-Sampling, Renewable Energy

Contents

1 Introduction **10**
- 1.1 **Problem Statement** . . . 10
- 1.2 **Research Gap** . . . 10
- 1.3 **Research Questions** . . . 11
- 1.4 **Scope and Scientific Contribution** . . . 11
- 1.5 **Thesis Structure** . . . 11

2 Methods **12**
- 2.1 **Energy System Optimisation** . . . 12
  - 2.1.1 **Problem Formulation** . . . 12
  - 2.1.2 **Time-Series** . . . 12
  - 2.1.3 **Energy and Power Based Formulation** . . . 13
- 2.2 **Down-Sampling: A Literature Overview** . . . 13
  - 2.2.1 **Down-Sampling in Energy System Optimisation** . . . 13
  - 2.2.2 **Alternative Down-Sampling** . . . 14
- 2.3 **Down-Sampling: Proposed Methods** . . . 14
  - 2.3.1 **Benchmark** . . . 15
  - 2.3.2 **Basic Statistics** . . . 15
  - 2.3.3 **Hybrid** . . . 15
  - 2.3.4 **Initial Representation Analysis** . . . 16
- 2.4 **Evaluation and Validation** . . . 17
  - 2.4.1 **Evaluation Metrics** . . . 17
  - 2.4.2 **Quality Check** . . . 18
  - 2.4.3 **Validation** . . . 18

3 Northwestern Europe Case Study **18**
- 3.1 **General Structure and Mathematical Formulation** . . . 18
- 3.2 **Input Data and Model Configurations** . . . 19
- 3.3 **Down-Sampling Experiments** . . . 20

4 Results **21**
- 4.1 **Initial Representation Analysis** . . . 21
  - 4.1.1 **Basic Statistics** . . . 21
  - 4.1.2 **Hybrid** . . . 23
  - 4.1.3 **Conclusion Initial Representation Analysis** . . . 24
- 4.2 **Model Results** . . . 24
  - 4.2.1 **Basic Statistics** . . . 24
  - 4.2.2 **Hybrid** . . . 29
  - 4.2.3 **Conclusion Model Results** . . . 33
- 4.3 **Quality Check** . . . 36
  - 4.3.1 **Investment Models** . . . 36
  - 4.3.2 **Operational Models** . . . 37
  - 4.3.3 **Conclusion Quality Check** . . . 38
- 4.4 **Validation** . . . 39
  - 4.4.1 **Basic Statistics** . . . 39
  - 4.4.2 **Hybrid** . . . 41
  - 4.4.3 **Conclusion Validation** . . . 44

5 Discussion **46**

6 Conclusion **49**

7 Acknowledgements **50**

Appendix A Algorithms Down-Sampling **55**
55appendix.B
Appendix B.1 **Overview of Sets, Parameters and Variables** . . . 55
Appendix B.2 **Objective Function** . . . 56
Appendix B.3 **Balancing Constraints for Consumer Energy Assets** . . . 57
Appendix B.4 **Balancing Constraints for Storage Energy Assets** . . . 57
Appendix B.5 **Balancing Constraints for Hub Energy Assets** . . . 57
Appendix B.6 **Balancing Constraints for Conversion Energy Assets** . . . 57
Appendix B.7 **Capacity Constraints: Maximum Output Flows Limit** . . . 57
Appendix B.8 **Capacity Constraints: Maximum Input Flows Limit** . . . 57
Appendix B.9 **Capacity Constraints: Lower Bound on Flows** . . . 57
Appendix B.10 **Transport Flow Capacity: Maximum Transport Flow Limit** . . . 57
Appendix B.11 **Transport Flow Capacity: Minimum Transport Flow Limit** . . . 57
Appendix B.12 **Extra Constraints for Energy Storage Assets: Maximum Storage Level Limit** . . . 57
Appendix B.13 **Extra Constraints for Energy Storage Assets: Cycling Constraints** . . . 58
Appendix B.14 **Investment Constraints: Maximum Investment Limit for Assets** . . . 58
Appendix B.15 **Investment Constraints: Maximum Investment Limit for Flows** . . . 58

Appendix C Input Data NWEU Case Study **58**

Appendix D List of Basic Statistics Experiments **59**

Appendix E Threshold Range and GridSearch Hybrid Down-Sampling **59**

Appendix F List of Hybrid Experiments **61**

Appendix G Thresholds Hybrid Experiments **62**

Appendix H Visual Overview Methods Quality Check **62**

Appendix I Results of Euclidean Distance and Clustering Analysis **63**

Appendix J Initial Representation Analysis **64**
Appendix J.1 **Energy Demand Profile** . . . 64
Appendix J.1.1 **Minima and Maxima** . . . 64
Appendix J.2 **Solar Profile** . . . 66
Appendix J.2.1 **Minima and Maxima** . . . 66
Appendix J.3 **Wind Onshore Profile** . . . 67
Appendix J.3.1 **Minima and Maxima** . . . 67
Appendix J.3.2 **Generation and Ramp Duration Curves** . . . 68
Appendix J.4 **Wind Offshore Profile** . . . 69
Appendix J.4.1 **Minima and Maxima** . . . 69
Appendix J.4.2 **Generation and Ramp Duration Curves** . . . 70

Appendix K Flow Breakdown Operational Models 2008 **71**

Appendix L Investment Decisions Models 2008 **75**
Appendix L.1 **Hourly Reference Models** . . . 75
Appendix L.2 **Basic Statistics** . . . 75
Appendix L.3 **Hybrid** . . . 76
Appendix L.4 **Initial Representation Analysis 2011** . . . 76
Appendix L.5 **Initial Representation Analysis 1997** . . . 79

List of Acronyms

RES Renewable Energy Resources

ESO Energy System Optimisation

TSA Time Series Aggregation

DS Down-Sampling

NWEU Northwestern Europe

EB Energy Based

PB Power Based

CAQE Cumulative Absolute Quantile Error

CSV Comma-Separated Values

NL The Netherlands

BE Belgium

GE Germany

FR France

UK The United Kingdom

IE Ireland

DK Denmark

CH Switzerland

AT Austria

TYNDP 10-year Network Development Plan

OCGT Open-Cycle Gas Turbine

ENS Energy Not Served

List of Tables

1 Overview of input data used from TYNDP 19
2 Investment limits applied in the investment model configurations 20
3 Overview of down-sampling experiments 20
B.4 Overview of all sets used in the Tulipa formulation 56
B.5 Overview of all parameters used in the Tulipa formulation 56
B.6 Overview of all variables used in the Tulipa formulation 56
C.7 Overview of all input CSV files required to build a case study with TulipaEnergyModel tool . . 58
D.8 List of experiments performed with benchmark and basic statistics down-sampling 59
F.9 List of hybrid down-sampling experiments 61
G.10 Thresholds used for hybrid experiments 62
I.11 Euclidean distances of all climate years from 1982 to 2016 compared to 2008 64

List of Figures

1 Energy blocks or average energy levels in energy-based formulation (left) and piece-wise linear instantaneous power profiles as represented in power-based formulation (right), reproduced from [1]. 13

2 Visual illustration of the general DS procedure for an hourly time series with 12 time steps to 3-hourly resolution. The original time series is divided into $12/3 = 4$ blocks, each containing 3 different data points (indicated in red, green and orange). To down-sample the blocks, a representative value is picked for the three data points in each block. The down-sampled time series has 12-time steps, with the representative value repeated 3 times to represent the 3 distinct data points in the blocks giving an energy level type appearance. 15

3 Visual illustration of the hybrid DS procedure for an hourly time series with 12 time steps to 3-hourly resolution. The time series is divided into $12/3 = 4$ blocks, each containing 3 different data points. To down-sample the blocks, either averaging or the maximum is used to represent three data points in each block based on signal variability. The down-sampled time series has 12-time steps, with the representative value repeated 3 times to represent the 3 distinct data points in the blocks giving an energy level type appearance. 16

4 Visual overview of the structure of the TulipaEnergyModel tool. The input data includes asset and flow information, availability profiles and desired resolution to run the model on. Tulipa itself contains the formulation and constraints and depending on the model configuration and input data different case studies can be built. 19

5 Overview of the countries in the Northwestern European case study, indicated in yellow. 19

6 Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 21

7 Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 22

8 Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods using different thresholds, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 23

9 Cumulative absolute quantile error of the ramp distribution in the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods using different thresholds, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 24

10 Overview of the results of the basic statistics down-sampling experiments in the investment model. On the x-axis the evaluation metrics are displayed including the objective value (total system costs), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. As opposed to the other metrics in the matrix the flow and net position behaviour indicated in this graph is the behaviour in the model after reallocation to 1 hourly level. 25

11 Overview of the results of the basic statistics down-sampling experiments in the investment model with increased limits. On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model, and the flow and net position behaviour in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value, investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2. 27

12 Overview of the results of the basic statistics down-sampling experiments in the operational model. On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the flows and net position. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not. 28
13 Overview of the results of the hybrid down-sampling experiments in the investment model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2. 29
14 Overview of the results of the hybrid down-sampling experiments in the increased limits investment model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value (total system cost), investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2. 31
15 Overview of the results of the hybrid DS experiments in the operational model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The thresholds are ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time, flows and net position. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not. 32
16 An overview of the energy not served after reallocation of the investment models (left) and investment model with increased limits (right) using the basic statistics methods. The energy not served is compared to that in the hourly reference model. The methods used for down-sampling are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions. 36
17 An overview of the energy not served after reallocation of the investment models (top) and investment model with increased limits (bottom) using the hybrid method. The energy not served is compared to that in the hourly reference model. The methods used for down-sampling are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions. 37
18 An overview of the energy not served after reallocation of the operational models with the basic statistics methods (top) and hybrid method (bottom). The energy not served is compared to that in the hourly reference model. The methods used for DS are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions. 38

19 Overview of the results of the validation of the basic statistics methods in the investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not. 39
20 Overview of the results of the validation of the basic statistics methods in the increased limit investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not. 40
21 Overview of the results of the validation of the basic statistics methods in the operational model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time, flows and net position. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value flows, and net position the differences are absolute while for the solve time, they are not. 41
22 Overview of the results of the validation of the hybrid methods in the investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not. 42
23 Overview of the results of the validation of the hybrid methods in the investment model with increased limits with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not. . . 43
24 Overview of the results of the validation of the hybrid methods in the operational model with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not. 44
E.25 Results of the grid search for the thresholds associated with the optimal CAQE ramps for the energy demand and solar profiles. 60
E.26 Results of the grid search for the thresholds associated with the optimal CAQE ramps for the wind onshore and wind offshore profiles. 61
H.27 Schematic overview of the methods for the quality check for the investment models (top) and operation models (bottom). 63

I.28 The silhouette score (left) of the k-means clustering, shows that the optimal cluster number is 2 (highest score) (left). The visualisation of k-means clustering with PCA (right), shows that the reference year 2008 and the two most different years (1887 and 1991) using Euclidean distance are in the same clusters. Thus the most different year from 2008 in terms of Euclidean distance in the third cluster was picked, which is 2011 (right). 63
J.29 Locations of minima and maxima in the energy demand profile compared to the original Dutch energy demand profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution. 65
J.30 Locations of minima and maxima in the energy demand profile compared to the original Dutch energy demand profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution. 66
J.31 Locations of minima and maxima in the solar profile compared to the original Dutch solar profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution. 66
J.32 Locations of minima and maxima in the solar profile compared to the original Dutch solar profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution. . . . 66
J.33 Locations of minima and maxima in the wind onshore profile compared to the original Dutch wind onshore profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution. 67
J.34 Locations of minima and maxima in the wind onshore profile compared to the original Dutch wind onshore profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution. 67
J.35 The generation and ramp duration curves of wind onshore profiles down-sampled with basic statistics methods to 2, 3 and 4-hourly resolution. 68
J.36 The generation and ramp duration curves of wind onshore profiles down-sampled with the hybrid method to 2, 3 and 4-hourly resolution. 68
J.37 Locations of minima and maxima in the wind offshore profile compared to the original Dutch wind offshore profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution. 69
J.38 Locations of minima and maxima in the wind offshore profile compared to the original Dutch wind offshore profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution. 69
J.39 The generation and ramp duration curves of wind offshore profiles down-sampled with the basic statistics methods to 2, 3 and 4-hourly resolution. 70
J.40 The generation and ramp duration curves of wind offshore profiles down-sampled with the hybrid method to 2, 3 and 4-hourly resolution. 70
K.41 A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the benchmark method on 2, 3 and 4-hourly resolution. 71
K.42 A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the minimum method on 2, 3 and 4-hourly resolution. 72
K.43 A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the hybrid method on the energy demand profile with a threshold of 0.04 on 2, 3 and 4-hourly resolution. 73
K.44 A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the hybrid method on the solar profile with a threshold of 0.06 on 2, 3 and 4-hourly resolution. 74
L.45 Investment decisions of the hourly reference investment model (left) and of the hourly reference investment model with increased limits (right). 75
L.46 Investment decisions in the investment model with basic statistics down-sampling experiments. 75
L.47 Investment decisions in the increased limits investment model with basic statistics down-sampling experiments 75
L.48 Investment decisions in the investment model (left) and in the investment model with increased limits (right) with the hybrid method experiments. 76
L.49 Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 76

L.50 Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 77
L.51 Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 77
L.52 Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid method, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 78
L.53 Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 79
L.54 Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 79
L.55 Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 80
L.56 Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid method, to 2, 3 and 4-hourly resolution as indicated on the x-axis. 80

1. Introduction

1.1. *Problem Statement*

Driven by societies' concern for the environment there is a growing penetration of renewable energy sources (RES) in power systems [2, 3]. These, RES such as solar and wind energy are subject to variable weather conditions and with the increased adoption of such technologies, uncertainty increases, leading to operational difficulties [4, 5]. Therefore a key priority is to achieve higher flexibility to accommodate the increasing levels of intermittent electricity generation [1, 3]. The flexibility of a power system is measured by how well it balances fluctuating demand and generation [1]. To offset the demand that is not met by RES an appropriate capacity of system resources needs to be present [1, 5]. Hence, to anticipate and allocate the operation of these supporting system resources accurate modelling is required to ensure both dependable and cost-effective electricity generation [4, 6].

A way to model this is by using energy system optimisation (ESO). In such models, the concept of unit commitment is used, either in the short-term with day-ahead unit commitment [1, 7, 8], or in the long-term with capacity or generation expansion planning [9, 10, 11, 12]. Day-ahead unit commitment is generally used to allocate the supporting system resources cost-efficiently, ensuring the system and its units operate within safe technical boundaries [1]. Capacity or generation expansion planning aims to identify the optimal mix of generation technologies to ensure there is enough capacity to meet future demand [13, 9]. However the liberalisation and decentralisation of electricity markets together with the increasing volatility of electricity production due to RES integration, are significantly complicating the application of ESO [14, 15]. This is because the complexity of models increases with the rising proportion of variable RES [16], with the inclusion of storage technologies [17], and with the growing interconnectedness between different energy sectors [14].

1.2. *Research Gap*

Expansion planning models are typically associated with a longer period and consequently with more data. Therefore a strategy to manage this significant computational challenge can involve decreasing the time domain [4, 18]. Various time series aggregation (TSA) methods are available for reducing the time domain, they can be categorised into combining consecutive time steps, i.e. down-sampling (DS) and segmentation, and merging of periods, i.e. using time slices and period clustering [14, 15, 19]. Overall TSA is used to decrease the complexity of the models while trying to maintain the highest possible level of model accuracy. The most commonly used TSA method uses clustering to find a typical period, such as representative days, weeks or months to represent multiple periods [14, 15].

The method of DS is used less and, in most works that concern ESO uniform DS techniques are used, where multiple hours are aggregated into a single value by default averaging or in other words maintaining the energy value [14]. These uniform DS techniques have also been used for a long time in time series data mining and time series aggregation in other disciplines where it is often called Piece-wise Aggregate Approximation [20, 21, 22, 23]. When used in ESO research this approach is often primarily focused on implementation in ESO models, with a notable lack of thorough examination of how the decrease in model complexity balances against the loss in model accuracy. Existing references reflect that DS using averaging is not the most effective or efficient method for aggregating time series data in ESO [24, 14]. However, the fact that averaging is the default for aggregation in DS indicates that exploring alternative ways to aggregate methods may provide a more precise representation of time series data, potentially enhancing model performance while reducing model complexity. To illustrate, with averaging, the total energy content or energy value is always preserved, which is important for long-term energy balancing and planning [13]. On the other hand, with averaging, peaks and ramps can be smoothed out which results in incorrectly representing a time series if it has oscillating amplitudes [15, 25]. This can potentially lead to less accurate modelling of periods with high demand or rapid fluctuations.

Considering alternative strategies to averaging, such as using the maximum or minimum to maintain extreme values [20, 21, 23], using the median to preserve the general trends and filter noise and outliers, or using the first or last value of a time block to preserve chronological trends, could enhance the accuracy of time series representation in DS and thereby improve model accuracy. Moreover, a hybrid approach similar to what [20, 21] do in the aggregation of financial time series could be considered. The method adapts to the local properties of the time series by using combinations of the methods mentioned above providing a better balance between preserving different characteristics of the data. Alternative DS techniques may provide a promising solution in ESO especially when also considering, that clustering methods can be time-consuming, which can result in a trade-off between the computational burden of clustering and the conservation of computational resources when using aggregated models [14].

1.3. ***Research Questions***

Following from the research gap outlined above, the main research question is:

'How effective are alternative DS methods in lowering the temporal resolution in ESO to reduce complexity while approximating accuracy on hourly resolution?'

This main research question is answered by addressing the research sub-questions below:

1. *How accurately can alternative down-sampling methods represent input time series in ESO models?*
2. *What is the impact of using the down-sampled input time series on the accuracy of ESO models?*
3. *What is the impact of using down-sampled input time series on the computational efficiency of ESO models?*

1.4. ***Scope and Scientific Contribution***

The research questions are addressed within the context of a Northwestern Europe (NWEU) case study. The effect of DS is evaluated with different model metrics, including the total system costs, investment decisions, flows, net position behaviour and solve time. The experiments were performed in investment and operational model configurations, and although the methods are applied in these specific contexts they are designed to be adaptable and relevant for broader applications. For example, they can be applied in similar models but on a larger scale or in other geographical regions. In addition, the findings can be used in applications where DS is used in combination with other techniques, such as with rolling horizon [26] or when DS typical or representative periods [27, 28]. Moreover, the application of the alternative DS techniques is not exclusively applicable to ESO models, but also to other applications that rely on time series and could benefit from reducing complexity while approximating accuracy on an hourly level. This could be for example in time series forecasting models [29, 30], trend detection in climate data [31, 32] or in anomaly detection [33]. By addressing these research questions this study makes contributions to ESO, by showing the potential advantages of alternative DS techniques and highlighting their possibility of enhancing the implementation of ESO models. The findings contribute to the understanding of the relationship between time series representation and model performance and they add to the understanding of the balance between model complexity and model accuracy.

1.5. ***Thesis Structure***

The structure of the rest of this research is as follows. In Section 2, the relevant background and an overview of the literature concerning DS will be provided. Moreover, the proposed DS methods used to answer the research questions will be presented. Subsequently, in Section 3 the set-up of the NWEU case study will be explained. Finally, in Section 4 the results will be discussed and in Section 5 and 6, the discussion and conclusion will be presented, respectively.

2. Methods

In this section, the relevant literature to the research questions will be presented together with a literature overview of DS methods. Moreover, the proposed methods for alternative DS will be outlined followed by the methods for the initial representation analysis of the down-sampled time series. Finally, the steps that were taken to evaluate and validate the results are outlined.

2.1. *Energy System Optimisation*

2.1.1. *Problem Formulation*

A simple ESO problem is typically formulated to find the most cost-effective way to meet energy demand while satisfying a set of constraints [34]. Usually, this is done by minimising the objective function, which is dependent on decision variables and cost or price parameters. Let us consider an electricity production problem, which is usually a part of ESO problems, as formulated in Equation 1.

$$\min \sum_{t=1}^{T} \sum_{i=1}^{N} (C_i \cdot P_{i,t}) \tag{1}$$

Where T is the total number of time periods considered in the problem, N is the number of electricity generation sources or technologies, C_i is the cost per unit of electricity produced by source i, and $P_{i,t}$ is the power output of source i at time t.

In the objective function above the decision variables of the optimisation are the power outputs of each electricity generation source i at each period t $P_{i,t}$. In the case of, for example, a capacity or generation expansion planning model, another decision variable would be the capacity to be installed for each technology and in the case of a day ahead unit commitment model a decision variable would be whether the units will be operated or not. As mentioned, the objective function is subject to a set of constraints that ensure the solution is viable and meets all system requirements [34]. Generally, constraints are included that concern the demand satisfaction and renewable availability in a model [1, 9]. The demand satisfaction constraints ensure that the total electricity generated meets the demand at all time steps, see Equation 2. The renewable availability constraints ensure that for renewable sources the output is constrained by availability, which can vary with time, see Equation 3, respectively.

$$\sum_{i=1}^{N} P_{i,t} = D_t, \quad \forall t \tag{2}$$

$$P_{i,t} \leq A_{i,t} \cdot P_{i,t}^{\max}, \quad \forall i, \forall t \tag{3}$$

Where $P_{i,t}$ is the power output of unit i in time step t, D_t is the demand in time step t, t is the time step, i is the generation unit and $A_{i,t}$ is the renewable availability of unit i in time step t.

The outcome when solving the problem is a set of optimal decision variables and an optimal objective function value. The solution of the model is valid if all decision variables satisfy the constraints [34]. An optimisation problem can be solved using for example linear programming for problems with linear relationships and constraints, or mixed-integer linear programming if the problem includes integer variables e.g., on/off decisions for power plants in unit commitment problems. The simplified formulation presented earlier provides a basic framework for ESO however, for example, capacity or generation expansion planning models or day-ahead unit commitment models may involve more complex formulations, such as ramping, capacity limit and balancing constraints.

2.1.2. *Time-Series*

In an ESO problem as previously described time series data can occur in various ways, e.g. they can directly influence the objective function, as seen with hourly electricity prices, but they can also enter the constraints, such as with the availability of RES or with electricity demand, as described by [34] and [5]. The computational time of solving the problem is dependent on the resolution associated with the used time series [16, 35]. This is because the complexity of the model is dependent on the number of variables and constraints, and this increases when using high resolutions [16, 35]. A higher resolution, however, enhances the accuracy of the solution of the model and lowering this resolution can possibly lead to infeasible model solutions [34]. This highlights the trade-off between reducing temporal resolution and decreasing computational complexity. Nonetheless, considering the EB formulation discussed above, through the application of alternative DS techniques, it may be feasible to adjust time series at a reduced temporal resolution in a way that preserves more information. Accordingly, in Section, 2.2 DS methods used in ESO and possible alternative DS methods that may be able to manage the trade-off described above will be discussed.

2.1.3. *Energy and Power Based Formulation*

The modelling of unit commitment problems can be done with two different approaches, namely with the conventional energy-based (EB) formulation and with the power-based (PB) formulation. In EB formulation energy production and consumption are represented as average values across scheduling periods that usually span one hour [1], see Figure 1. EB models often fail to deal with ramp capabilities, which arise from applying ramp constraints to the average energy levels [36]. Conversely, in PB formulation a clear distinction is made between energy and power, and demand and generation are modelled as hourly piece-wise linear functions representing instantaneous power trajectories [9], see Figure 1. Here system flexibility is exploited by accurately incorporating ramping limits and operational reserves to ensure the anticipated and actual flexibility from the generation assets is achieved. This is possible because of the distinction between power and energy in the PB formulation [9]. PB models use investment more efficiently because they represent the reality of flexibility requirements better and they adequately exploit flexibility capabilities in the system [9]. With the piece-wise linear power trajectories, the variability of demand and generation profiles is represented better than with the averaged energy levels used in the EB formulation. Flexibility capabilities and system requirements are therefore represented more accurately. This is in contrast with the step-wise energy functions used in EB models, which may overestimate the flexibility of the system. In addition to this gain in more realistically capturing flexibility capabilities, with PB modelling a gain can be made computationally and the solution of the models improves [1, 9, 37].

The piece-wise linear representation of power trajectories is an improvement compared to the step-wise energy functions in EB models. However, implementing DS to reduce temporal resolution could be beneficial in both EB and PB formulations, as in EB models information on variability within and between time steps is lost due to averaging them in hourly energy blocks and in PB models errors are introduced in the energy content because in the formulation it is still assumed that power changes linearly within the hour [1]. Accordingly, through the application of DS techniques from other fields, it may be feasible to also adjust power profiles at a reduced temporal resolution in a manner that preserves more information regarding ramps and peaks.

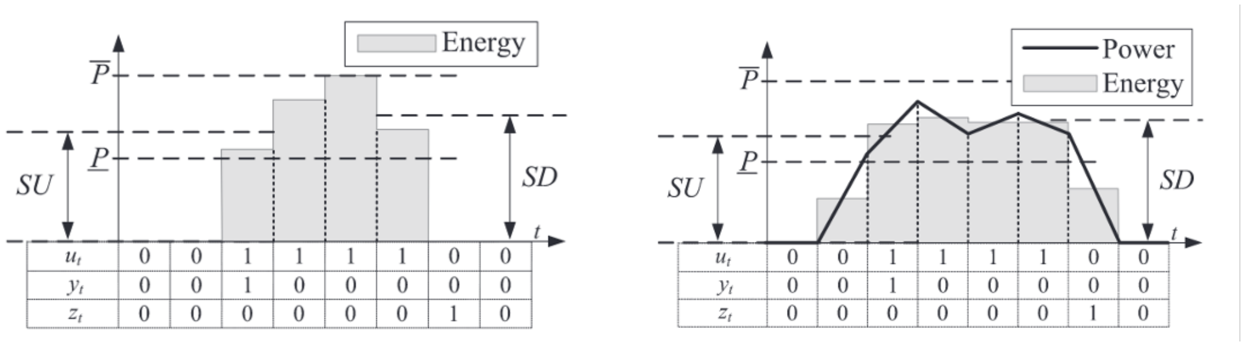


Figure 1: Energy blocks or average energy levels in energy-based formulation (left) and piece-wise linear instantaneous power profiles as represented in power-based formulation (right), reproduced from [1].

2.2. *Down-Sampling: A Literature Overview*

TSA is an umbrella term for methods that attempt to lower the resolution of time series data. In ESO models these methods are used to reduce the computational burden while maintaining accuracy in the representation of data characteristics and model solutions [4, 38]. TSA methods include merging adjacent steps or DS, using gradients for segmentation [26], and identifying representative periods to capture periodic patterns [14]. As described by Hoffman et al., [14] DS is not used often compared to using for example representative periods in the field of ESO. However, the method is still regarded to have potential, therefore in the following sections, an overview will be provided of the use of DS techniques within ESO and their use across other fields.

2.2.1. *Down-Sampling in Energy System Optimisation*

When DS, time series are generally reduced by aggregating consecutive time steps into one representative time step [14]. To illustrate, a time series of 8760 time steps over one year, can be reduced to a time series of 2190 time steps by aggregating every four time steps into one [24]. In the field of ESO, this aggregation is usually done by taking the average or the energy value of the consecutive time steps [14].

For example, [39, 38, 24] use this averaging DS method as a way to manage the temporal resolution in ESO models. In [24] they down-sample 3,6,12 or 24 consecutive hours to reduce the computational time and they show that this reduction is at the cost of accurately capturing fluctuations in RES. Similarly, in [39] and [38] DS is applied using averaging to reduce time series from hourly to 2 or 4-hourly intervals, to improve computational efficiency. The results of these works indicate that while these approaches offer computational improvement there is a trade-off with approximating the accuracy at an hourly level, especially in representing the dynamics of RES. Similarly, [40, 41, 42] examine the effect of sub-hourly resolutions on energy system modelling. They highlight that with these sub-hourly resolutions, it is possible to capture more variability and thereby a more accurate representation of system operation. They suggest that these higher temporal resolutions can be used to improve system flexibility and reliability [41]. It is also pointed out however that these methods substantially increase the computational demand, again emphasising the need for a balanced approach when adjusting temporal resolution. Other research uses more sophisticated methods, for instance, [26] use a so-called rolling horizon strategy in combination with DS. With this method, the model timeline is segmented, and DS is applied to segments of the time series that are further in future to reduce computational complexity while ensuring a detailed analysis in the short term. Furthermore, [27] and [28] also explore combined approaches, such as DS of typical periods.

Overall, the literature reflects that DS alone is not the most used method and may not be the most effective or efficient method for aggregating time series [24]. However, the reliance on mostly averaging when DS suggests an unexplored area where alternative ways to DS within ESO may offer a more accurate representation of time series data leading to improved model performance.

2.2.2. ***Alternative Down-Sampling***

As discussed above for ESO, averaging is most often used when DS is applied to aggregate time steps [14]. In other disciplines for example in the representation of time series data, in data mining and anomaly detection, DS using averaging often referred to as Piece-wise Aggregation Approximation, has also been used often [20, 21, 22]. However, next to averaging other aggregation methods can be explored as well, for example by using another statistical metric to find a representative value for aggregation [43, 21]. The intuition here is that different basic statistics methods can capture different characteristics of a time series. For example in the case of an energy profile, with averaging the overall energy content is always preserved when DS. However, this approach can smooth out fluctuations in the energy profile [20, 21, 25], potentially leading to less accurate representation and modelling of peak demand periods and rapid changes in energy demand. Moreover, when accurately representing an electricity system in ESO modelling preserving extreme values, and ramps is also critical [44]. This is because peaks in the time series correspond to periods of high energy demand and accurately capturing them is necessary to ensure that an optimisation model can plan for an adequate energy supply [44]. Additionally, ramps indicate rapid changes in energy demand or supply and must be preserved in a down-sampled time series to allow the optimisation model to respond effectively to sudden changes. Thus, alternative strategies to averaging can be using the maximum to preserve the peak values or using the minimum to preserve minimum trends, like [20] do in combination with averaging to aggregate financial time series. A similar method to averaging could be using the median which filters noise and outliers while preserving the overall trend while less sensitive to outliers which can prevent skewing the distribution of, for example, energy values in a time series. Finally, the last or first value can be used as a representative value for a time block upon aggregation, maintaining chronology, for example, the onset or end of a ramp.

Moreover, a combination of the above-mentioned sampling strategies could be employed as well. For example, similar to what is done in adaptive trace segmentation-based DS one could take into account the derivative activity or gradients of time steps in the time window that will be aggregated to determine whether it may be slow-changing or fast-changing parts of the signal [45, 20, 21, 25]. Depending on this information a different sampling strategy could be picked instead of applying the same metric or strategy for all the time windows. This makes the DS process more adaptive to the local characteristics of the time series, compared to smoothing out important information with averaging [20, 25, 21]. Such a hybrid method can address these issues by dynamically choosing between metrics for aggregation to balance the need to preserve the overall energy content with for example the accurate representation of peaks and ramps.

2.3. ***Down-Sampling: Proposed Methods***

Different DS strategies were employed to lower the temporal resolution of a time series and the general procedure remained consistent across the different DS methods, see Algorithm 1 in Appendix A. When DS an hourly time series $\{(t_i, y_i)\}_{i=1}^{M}$ with M time steps to an N-hourly time series, the time series was divided into $k = M/N$ blocks, with each block containing N data points. Then, a specific sampling strategy was applied to down-sample the data in each block into one value representing the N different data points. The resulting down-sampled time series still has M time steps, but with the representative value of each block repeated N times giving an energy level step-wise type function, see Figure 2 where averaging is used as a sampling strategy. The particular sampling strategy varies depending on the DS method used. In the next sections, the specific strategies will be presented.

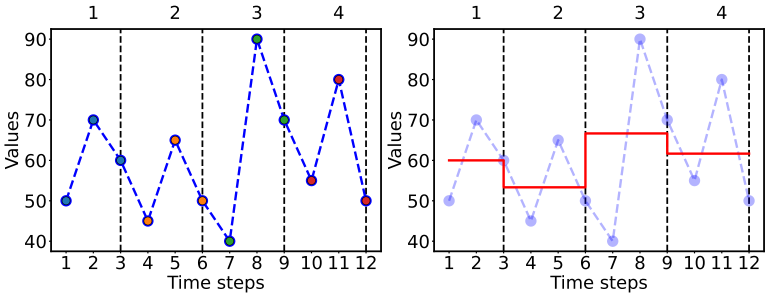


Figure 2: Visual illustration of the general DS procedure for an hourly time series with 12 time steps to 3-hourly resolution. The original time series is divided into 12/3 = 4 blocks, each containing 3 different data points (indicated in red, green and orange). To down-sample the blocks, a representative value is picked for the three data points in each block. The down-sampled time series has 12-time steps, with the representative value repeated 3 times to represent the 3 distinct data points in the blocks giving an energy level type appearance.

2.3.1. ***Benchmark***

All DS methods were compared to using the benchmark method, where averaging was used to find a representative data point for each N-block and to lower the resolution of the time series, see Equation 4.

$$y_{ds} = \frac{1}{N} \sum_{j=i}^{i+N-1} y_j \tag{4}$$

Where, y_{ds} is the down-sampled representative value, i is the time step, and y_j is the value in time step j.

2.3.2. ***Basic Statistics***

For the DS methods using other basic statistics sampling strategies, the maximum, minimum, median, first or last value was used to find a representative data point in each N-block. Additionally, linear interpolation was used, see Equation 5 – 6. For each block, the midpoint was calculated using Equation 5 and subsequently, the down-sampled value for that block was calculated with Equation 6.

$$t_{mid} = \frac{t_i + t_{i+N-1}}{2} \tag{5}$$

$$y_{ds} = y_i + \frac{y_{i+N-1} - y_i}{t_{i+N-1} - t_i} \cdot (t_{mid} - t_i) \tag{6}$$

Where t_{mid} is the midpoint, t_i and t_{i+N-1} are the time steps at the beginning and end of the block, respectively, and y_i and y_{i+N-1} are the values at the time steps of the beginning and end of the block, respectively.

2.3.3. ***Hybrid***

DS using the benchmark method preserves the overall energy content, however, it smooths out significant peaks and ramps, potentially leading to less accurate modelling of high-demand periods and rapid changes [20]. We used the hybrid method to address these concerns by choosing between the benchmark averaging and maximum values based on a standard deviation threshold within each block of data, see Figure 3. This way the method balances preserving the energy content by using the average and preserving peak information by using the maximum. Thus, depending on the threshold, when the variability is high in a block the maximum is used and when it is low the average benchmark is used. The threshold for the standard deviation varies with the characteristics of the time series being down-sampled. It ranges between two extremes, where at the one end the threshold is set such that the maximum value is always chosen, i.e. a low value such that the standard deviation in the block is always higher, see Algorithm 2. At the other end, the threshold is set such that the benchmark value is always chosen, i.e. a high value such that the standard deviation in the block is always lower than the threshold, see Algorithm 2 in Appendix A. We explored different values within this range. Additionally, a grid search was performed to find the threshold that provides the best representation of the time series in terms of capturing the ramps. The different thresholds used will be specified in the experimental setup in Section 3.

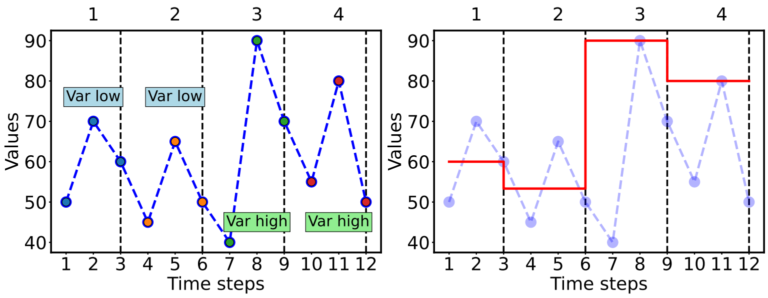


Figure 3: Visual illustration of the hybrid DS procedure for an hourly time series with 12 time steps to 3-hourly resolution. The time series is divided into 12/3 = 4 blocks, each containing 3 different data points. To down-sample the blocks, either averaging or the maximum is used to represent three data points in each block based on signal variability. The down-sampled time series has 12-time steps, with the representative value repeated 3 times to represent the 3 distinct data points in the blocks giving an energy level type appearance.

2.3.4. ***Initial Representation Analysis***

The initial representation analysis serves as a way to gain an initial understanding of how the down-sampled time series represents the original data. Although the measurements do not provide precise information on the impact of DS on the accuracy of an ESO model, they do help to identify characteristics that may play a role in minimising the decrease in model accuracy. The Cumulative Absolute Quantile Error (CAQE) was used to indicate the errors in the representation of the distribution of energy values. The CAQE measures the differences between the down-sampled data quantiles with the original data quantiles. The quantiles were determined using Equation 7 and the percentiles in Equation 8 and subsequently the CAQE was calculated using Equation 9.

$$\mathrm{Q}_{\text{time-series}} = \{\mathrm{Q}(y, p_i) \mid i = 1, 2, \ldots, 49\} \tag{7}$$

$$\text{percentiles} = [p_1, p_2, \ldots, p_{49}] \tag{8}$$

Where y represents the values of the time series, and $\mathrm{Q}(y, p_i)$ is the value at the i-th percentile of y. The percentiles that were used in the calculation are linearly spaced between 0.02 and 0.98, see Equation 8

$$\text{CAQE} = \sum \left|\mathrm{Q}_{\text{reference}} - \mathrm{Q}_{\text{down-sampled}}\right| \times \text{level} \tag{9}$$

Where the level is a scaling factor depending on the aggregation level. The results of this analysis are presented in Section 4.1.

The maxima and minima of a 48-time-step sample of the down-sampled time series were compared to the corresponding 48-time-step sample of the original time series. This was done to investigate the representation of extreme values. For the results of this analysis, see Section Appendix J. To evaluate the ramp representation of the DS methods, the Cumulative Absolute Quantile Error (CAQE) of the ramps was calculated, using Equation 10-12. First, the ramps of the series were calculated using Equation 10 and the ramp quantiles were calculated using Equation 11. Finally, the CAQE of the ramps was calculated with Equation 12.

$$\text{ramp}_i = \frac{y_{i+1} - y_i}{\text{level}} \tag{10}$$

Where y_i and y_{i+1} are consecutive data points in the time series and level is the aggregation level or resolution.

$$\text{quantiles}_{\text{ramp, reference}} = \text{quantile}(\{\text{ramp}\}, p_i) \tag{11}$$

Where p represents the set of percentiles, as defined in Equation 8.

$$\text{CAQE}_{\text{ramps}} = \sum \left|\mathrm{Q}_{\text{ramps, reference}} - \mathrm{Q}_{\text{ramps, ds}}\right| \times \text{level} \tag{12}$$

Where the level is a scaling factor depending on the aggregation level or resolution. The results of this analysis are presented in Section 4.1. It must be noted that the mathematical formulation of the models used in the case study does not contain ramp constraints, see Appendix B. Nonetheless, the representation of ramps in the down-sampled profiles is regarded as important as it reflects a more realistic version of the original profile.

2.4. Evaluation and Validation

This section will discuss the steps taken to evaluate and validate the impact of DS on the model results. First, the computation of the evaluation metrics will be presented. Subsequently, the methodological steps for quality checks of the models will be discussed and finally, the validation of the experiments will discussed.

2.4.1. Evaluation Metrics

For the operational model configuration, the objective value or total system costs, including the total operational costs and investment costs, flows, and net position were evaluated. For the investment models, the metrics concern the objective value, the investment decisions, the flows and net position behaviour after reallocation to hourly resolution, see Section 2.4.2. All metrics were compared to the hourly reference model, as well as to benchmark model results at different resolutions. Next to these metrics, the % difference of the model solve time was evaluated, i.e. the solve time of the down-sampled models compared to the solve time of the reference model on 1 hourly resolution, see Equation 13. Please note that all metrics are in absolute terms, except for the solve time.

$$\%\text{ Diff Solve Time} = \frac{\text{Solve Time}_{\text{ds-model}} - \text{Solve Time}_{\text{reference}}}{\text{Solve Time}_{\text{reference}}} \times 100\% \tag{13}$$

The objective value was extracted from the model output and the absolute % difference with the objective value of the hourly reference model was calculated to give an indication of the error in the objective value, as in Equation 14.

$$\%\text{ Diff ObjValue} = \frac{|\text{ObjValue}_{\text{ds-model}} - \text{ObjValue}_{\text{reference}}|}{\text{ObjValue}_{\text{reference}}} \times 100\% \tag{14}$$

To represent the difference in flows, the mean absolute error (MAE) between the asset flows in the experiments and the hourly reference model was calculated. To normalise this value the difference was divided by the mean of the asset flows in the hourly reference model, see Equations 15-17. The flows (MWh) represent the movement of energy between the physical assets that consume, store, balance, or convert energy. For each generation in the flows in time step i for N is 8760 were considered.

$$\text{mean\_flows\_1h} = \frac{1}{N}\sum_{i=1}^{N} \text{asset\_flows\_1h}[i] \tag{15}$$

$$\text{MAE(asset\_flows)} = \frac{1}{N}\sum_{i=1}^{N} |\text{asset\_flows\_1h}[i] - \text{asset\_flows\_ds}[i]| \tag{16}$$

$$\text{Normalised MAE(asset\_flows)} = \frac{\text{MAE(asset\_flows)}}{\text{mean\_flows\_1h}} \times 100\% \tag{17}$$

To represent the difference in net position, the export and import flows were filtered out and they were summed for each country, see Equation 18. The percentage difference for the net position of each country compared to the net position of that country in the hourly reference model was determined using Equation 19. Finally, the total net position absolute % difference compared to the hourly reference was calculated by summing all absolute % differences of net position by country and dividing by the number of countries evaluated, see Equation 20.

$$\text{Net Position}_{\text{country}} = \sum_{i=1}^{N} \text{imp flows}_{\text{country}} + \sum_{j=1}^{M} \text{exp flows}_{\text{country}} \tag{18}$$

$$\%\text{ Diff Net Pos}_{\text{country}} = \frac{|\text{Net Pos}_{\text{country, orig1h}} - \text{Net Pos}_{\text{country, ds}}|}{\text{Net Pos}_{\text{country, orig1h}}} \times 100\% \tag{19}$$

$$\text{Total Net Pos \% Diff} = \frac{\sum_{country=1}^{C} (\%\text{ Diff NetPos}_{\text{country}})}{\sum_{country=1}^{C} \text{Country}} \tag{20}$$

The absolute % difference with the hourly reference for the investment decisions was calculated using Equation 21. The investment decisions include the total MW in capacity installed after solving the model. The investment decisions for battery, coal, gas, OCGT, solar, onshore wind and offshore wind were considered.

$$\%\text{ Diff Inv Dec} = \frac{|\text{Inv Dec}_{\text{ds-model}} - \text{Inv Dec}_{\text{reference}}|}{\text{Inv Dec}_{\text{reference}}} \times 100\% \tag{21}$$

2.4.2. *Quality Check*

A quality check for the models with down-sampled input profiles was performed to evaluate their performance after reallocation to hourly resolution. For the investment model designs, the investment decisions from the down-sampled model experiments were extracted and used as input in an hourly operational reference model. Subsequently, the energy not served (ENS) was evaluated to investigate how close the behaviour in balancing demand and generation is compared to the hourly reference. Moreover, the flows and net position of the down-sampled model after reallocation were evaluated. This was done specifically after the quality check, to separate the effect of differences in investment decisions and DS on the flows and net position. For the operational models, the operational decisions, i.e. the asset flows per time step, were extracted from the model output and used as input profiles in the hourly operational reference model to limit the generation capacity. After solving, the ENS was evaluated to see how the model behaves with the limited generation capacity. See Figure H.27 for a visual overview of the methods for the quality checks. The ENS metric serves as an indicator of the quality of the DS experiments, revealing how the models align generation with demand. Models that have a value of ENS that is closer to the hourly reference, match the behaviour of the hourly reference better.

2.4.3. *Validation*

To validate the results of the DS experiments, each experiment was repeated using two different climate years. For this analysis, the input profiles for solar, onshore wind, offshore wind, energy demand, run-of-river hydro, reservoir hydro, and pumped hydro availability were extracted for each climate year from 1982 to 2016 from the TYNDP data, as outlined in Table 1. Using Euclidean Distance, the two climate years most dissimilar to the reference year 2008 used in the primary experiments were determined. The analysis showed that 1997 and 1991 were the most distinct years from 2008, see Appendix I. To further validate that these selected years are not only different from the reference year but also distinct from each other, k-means clustering was used. The optimal number of clusters was determined using the Silhouette Score [46] and was found to be two, see Appendix I. However, since two years that are both different from the reference year 2008 and each other are needed, at least three clusters are required to validate the choice. The next highest silhouette score was three. The k-means clustering using an optimal cluster number of three revealed that the reference year 2008 was assigned to Cluster 2, while both years 1997 and 1991 were assigned to Cluster 1, see Appendix I. Consequently, the year 2011 was selected as it was the year most different from 2008 in Euclidean distance and belonged to Cluster 3. Section 4.4, presents the results of the validation with climate years 2011 and 1997. In Appendix L.4 and Appendix L.5 the results of the initial representation analysis of the profiles from the climate years 2011 and 1997 are summarised.

3. Northwestern Europe Case Study

In this section the set-up of the NWEU case study, including the general model structure, the collection and preprocessing of input data and the model configuration variations will be discussed. Moreover, an overview of the experiments performed will be presented.

3.1. *General Structure and Mathematical Formulation*

To build the NWEU case study version 0.7 of TulipaEnergyModel.jl was used[1]. TulipaEnergyModel.jl is a Julia-based optimisation tool designed for the electricity market and it uses the JuMP.jl package for its functionality. Tulipa operates on two primary concepts, energy assets and flows. Energy assets are the physical components involved in the energy system, that produce, consume, store, balance, or convert energy. Flows represent the movement of energy between assets. Tulipa allows for running different scenarios and energy problems. For the scenarios and energy problems built, Tulipa aims to find the best investment and operational strategies across a range of assets. It ensures a balance across these assets, considering flow constraints. The tool can accommodate hourly to multi-hour resolutions. Please refer to Appendix B for the mathematical formulation of Tulipa version 0.7. Users can construct detailed scenarios by providing input data in a specified format, as listed in Appendix C. Tulipa then builds the mathematical model, incorporating the predefined formulations and constraints, see Figure 4.

[1] https://tulipaenergy.github.io/TulipaEnergyModel.jl/v0.7/

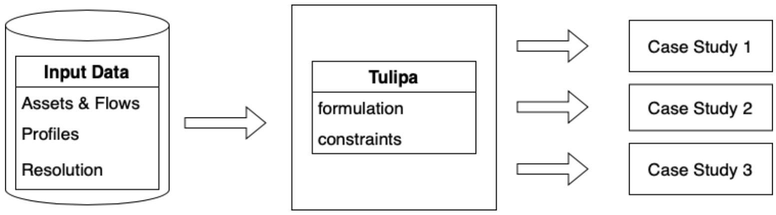


Figure 4: Visual overview of the structure of the TulipaEnergyModel tool. The input data includes asset and flow information, availability profiles and desired resolution to run the model on. Tulipa itself contains the formulation and constraints and depending on the model configuration and input data different case studies can be built.

3.2. Input Data and Model Configurations

The NWEU case study includes the Netherlands (NL), Belgium (BE), Germany (GE), France (FR), The United Kingdom (UK), Ireland (IE), Denmark (DK), Switzerland (CH) and Austria (AT), see Figure 5. The model uses a list of different CSV files as input, detailing the assets, flows, their profiles, partitions, and representative periods according to the Tulipa structure[2], see Figure 4. For a complete list and description of each type CSV file, please refer to Table C.7 in Appendix C.

The main data source that was used to set up these files is TYNDP using the 2030 scenario and 2008 for the climate year [47]. Table 1 provides a detailed summary of the files and information used from TYNDP, specifying their sources and applications. The different generation technologies that were considered are solar, wind onshore, wind offshore, nuclear, gas, hydro, coal and OCGT. The balancing technologies that were considered are OCGT, pumped hydro closed, pumped hydro open, hydro reservoir and battery.

Figure 5: Overview of the countries in the Northwestern European case study, indicated in yellow.

Table 1: Overview of input data used from TYNDP

File Name	**Contents & Purpose**
assets-data.csv	Capacity, Initial Capacity, Initial Storage Capacity, Initial Storage Level: *223010_Updated_Electricity_Modelling_Results.*
flows-data.csv	Capacity inter-country connection (transport) flows and peak demand: *223010_Updated_Electricity_Modelling_Results.*
profiles-rep-periods -availability.csv	Solar: *PECD_LFSolarPV_2030_edition 2021.3.* Offshore wind: *PECD_Offshore_2030_edition 2021.3.* Onshore wind: *PECD_Onshore_2030_edition 2021.3.*
profiles-rep-periods -demand.csv	Demand: *Demand_TimeSeries_2030_NationalTrends.*
profiles-rep-periods -inflows.csv	Hydro: *PEMMDB_XX00_Hydro Inflow_2030.*

Three different types of model configurations were employed, an operational model and two variations of an investment model. In the operational model, investment is disabled for all flows and assets, which is specified in

[2] https://tulipaenergy.github.io/TulipaEnergyModel.jl/dev/how-to-use/#input

the assets-data.csv and flows-data.csv files by setting the investable parameter to 'FALSE'. In the investment models, the greenfield condition was assumed, where all technologies and transmission lines are invested in and are constructed overnight from zero. Investment costs were annualised using the net present value, calculated with an annual discount factor of 0.04. The cost data were acquired from the Danish Energy Agency [48].

Table 2 summarises the investment limits used for the investment and increased limits investment model. For renewables limited by natural resources, the capacity from the TYNDP scenario results was used. For controllable technologies, a 1.2x capacity limit based on the TYNDP results was assumed. Nuclear and hydro plants were regarded as not investable because they are typically large projects planned several years in advance. The investable parameter was set to 'TRUE' in the assets-data.csv and flows-data.csv files for these models. All models were run in a Julia environment on a personal computer with a 3.3 GHz Dual-Core Intel Core i7 processor and 16 GB of memory. The optimisation problems were solved with Gurobi using linear simplex, and with the nodefile setting at 0.5.

Table 2: Investment limits applied in the investment model configurations

<table>
<tr><th>Technology</th><th>Capacity Investment Limit</th><th>Initial Capacity</th><th>Capacity Investment Limit (Increased)</th><th>Capacity (Adapted)</th></tr>
<tr><td>AC transmission lines</td><td rowspan="3">1x TYNDP result</td><td rowspan="7">0</td><td>1x TYNDP result</td><td rowspan="7">0</td></tr>
<tr><td>Solar</td><td rowspan="2">5x TYNDP result</td></tr>
<tr><td>Wind on/off-shore</td></tr>
<tr><td>Gas</td><td rowspan="4">1.2x TYNDP result</td><td rowspan="4">1.2x TYNDP result</td></tr>
<tr><td>OCGT</td></tr>
<tr><td>Coal</td></tr>
<tr><td>Battery</td></tr>
<tr><td>Nuclear</td><td>Not investable</td><td>1x TYNDP result</td><td>Not investable</td><td>1x TYNDP result</td></tr>
</table>

3.3. ***Down-Sampling Experiments***

The input profiles that were down-sampled include the solar, wind onshore, and demand, stored in the profiles-rep-periods-availability.csv and profiles-rep-periods-demand.csv files of the NWEU case study. The hydro inflow profiles in the profiles-rep-periods-inflows.csv file were excluded from DS because they were not available in hourly resolution in TYNDP. They were only available in daily and weekly resolutions, lacking the granularity required for this analysis. Each profile was down-sampled to 2, 3, and 4-hourly resolutions. The resolution of the profiles is specified in the assets-rep-periods-partitions.csv and flows-rep-periods-partitions.csv files. When the input profiles were down-sampled to an N-hourly resolution, the partition of all assets and flows was set to N, such that the complete model operated at the same resolution. The specification used for the partition of the profiles is 'uniform'[3]. To apply the hybrid method on the input profiles, threshold ranges were determined and additionally, an optimal threshold was identified with a grid search optimising for the CAQE ramps to get the optimal representation of ramps, see Appendix E. For the hybrid method only one profile was down-sampled at a time, while the others were down-sampled using the benchmark method, resulting in a variation of experiments, for each profile. Specifically, four experiments with different thresholds from the determined range and one experiment with the optimal threshold from the grid search. In all cases, a uniform resolution was maintained in the model. Please refer to Appendix E for the ranges of thresholds that were explored, together with the results of the grid search. In Table 3, the different experiments with the three different model configurations are listed, please refer to Appendix D and Appendix F for a complete list of all separate experiments.

Table 3: Overview of down-sampling experiments

Name	**Model**	**Resolutions**	**Down-Sampling**
op_reference	operation	1 hourly	x
inv_reference	investment	1 hourly	x
inv_increased_reference	investment increased	1 hourly	x
op_benchmark	operation	2,3,4 hourly	benchmark
inv_benchmark	investment	2,3,4 hourly	benchmark
inv_increased_benchmark	investment increased	2,3,4 hourly	benchmark
op_bs	operation	2,3,4 hourly	basic statistics
inv_bs	investment	2,3,4 hourly	basic statistics
inv_increased_bs	investment increased	2,3,4 hourly	basic statistics
op_hybrid	operation	2,3,4 hourly	hybrid
inv_hybrid	investment	2,3,4 hourly	hybrid
inv_increased_hybrid	investment increased	2,3,4 hourly	hybrid

[3]https://tulipaenergy.github.io/TulipaEnergyModel.jl/dev/how-to-use/#assets-rep-periods-partitions-definition

4. Results

This section presents the results of the DS experiments in the NWEU case study. First, the initial representation analysis of the DS methods will be discussed. Next, the model results of the DS experiments will be presented and finally, the results of the quality check and validation will be presented.

4.1. Initial Representation Analysis

4.1.1. Basic Statistics

Figure 6 illustrates the CAQE of the down-sampled profiles compared to the original profiles on different aggregation levels. From the four graphs it can be observed that upon decreasing resolution, the CAQE increases. This trend indicates that higher aggregation levels used in DS, result in a larger error in capturing the energy value distribution of the original profiles. The benchmark method (blue), consistently exhibits a low CAQE across all aggregation levels and profiles. Except for the maximum (green) and minimum (red) methods, most basic statistics methods perform similarly across the different profiles at 2-hourly resolution. The trend persists at the 3 and 4-hourly resolutions across all profile types, where the errors associated with the minimum and maximum methods remain substantially higher. This indicates that with the latter methods the error in energy value distribution introduced by DS is high. In contrast, at the 3 and 4-hourly resolutions, the median (orange), midpoint (pink), first (purple) and last (brown) methods have a lower CAQE than the benchmark in varying combinations for the different profiles. Moreover, for the solar, wind onshore and offshore profiles the midpoint, first and last methods have a lower CAQE than the benchmark at all resolutions. This indicates a more accurate representation of the energy value distribution. Lastly, the linear interpolation method (grey) scores similarly to the benchmark method, and for all profiles on the 4-hourly resolution, the CAQE is higher than the benchmark.

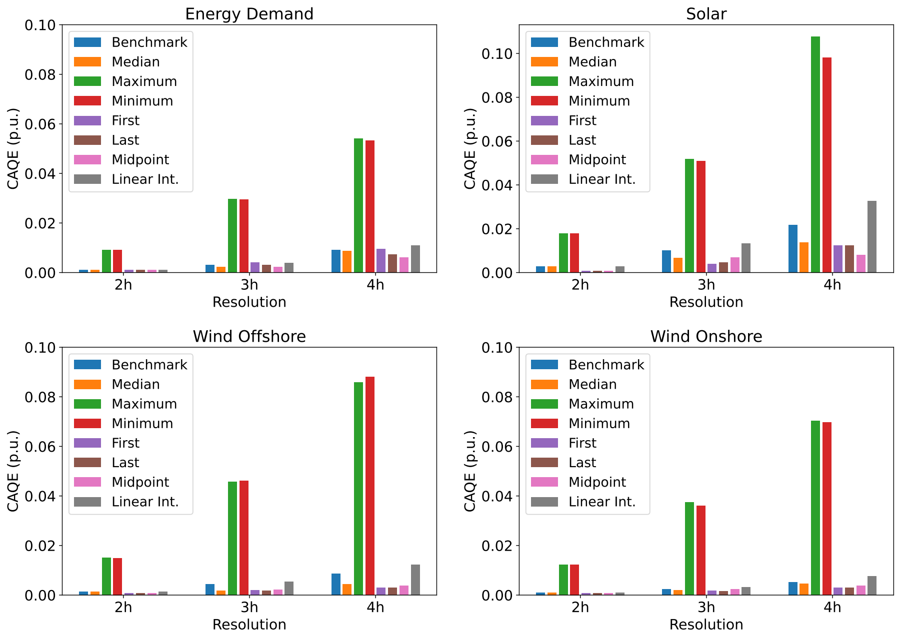


Figure 6: Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

The higher errors for the maximum and minimum methods occur because the extreme values are overestimated. This effect can be observed when examining the minima and maxima locations and the shape of the profiles after DS, see Appendix J.29-J.38. These figures show that while the minimum and maximum methods preserve extremes, the overall distance between the original profile and the down-sampled profile is large, especially at lower resolutions. Even for the solar profile, where the maximum method might be expected to perform better because of the daily peaks, a slight displacement of those peaks introduces an error in the energy value distribution. The lower CAQE of the linear interpolation method than the benchmark at 4-hourly resolution is likely because it assumes a linear relationship between two data points. This may not accurately capture the behaviour in energy demand, solar, and wind profiles. The lower CAQE for the median method is likely because it selects the middle value within each block and filters out extreme values. As can be seen in Appendix J.29-J.38, the down-sampled profiles using the median method follow the shape of the original profile

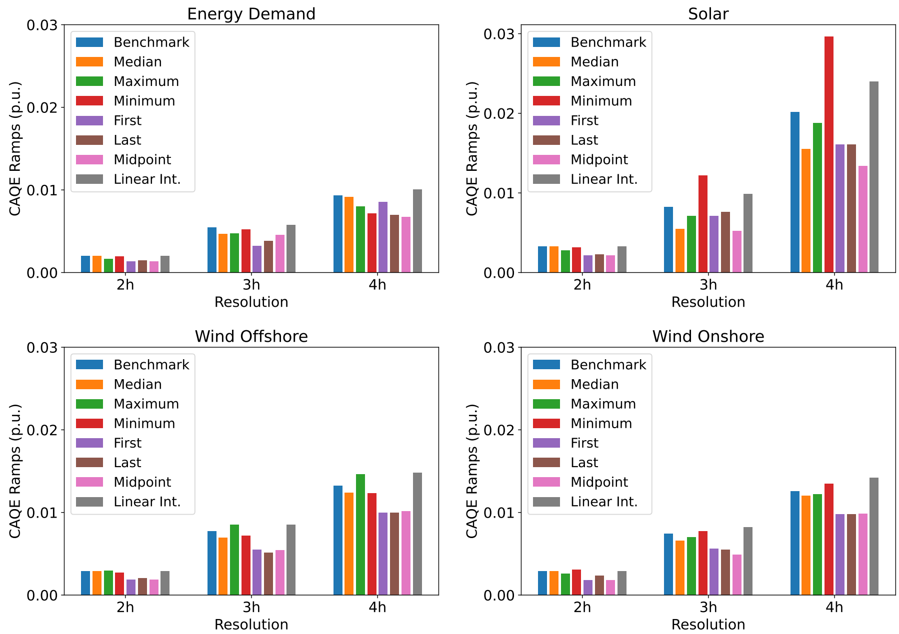


Figure 7: Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

slightly better. Although the median performs similarly to the benchmark, it sometimes captures minima and maxima locations slightly better. This is likely because taking the median value is less sensitive to outliers than averaging in the benchmark and therefore it is better at capturing the energy value distribution. The midpoint method shows similar performance to the median method and this is also reflected in the down-sampled profile shapes in Appendix J.29-J.38. For the solar and wind onshore and offshore profiles, the midpoint method has a lower CAQE than the benchmark across all resolutions. Even at the 2-hourly resolution, the midpoint method captures the overall profile shape and maxima more accurately.

For the solar, energy demand, wind onshore and wind offshore profiles, both the first and last methods have a lower CAQE than the benchmark at all resolutions. The profiles in Appendix J.29-J.38, show that these methods better capture maxima, particularly at the 2-hourly resolution. This is most pronounced for the solar and wind profiles. The last method, can sometimes better reflect the cumulative trend within the block, particularly in profiles where the final value might be more representative of the overall trend. These methods may outperform the benchmark in certain cases because the averaging can be sensitive to extreme values skewing the distribution of the data.

Figure 7 illustrates the CAQE of the ramps in the basic statistics experiments. Similar to the CAQE, the CAQE of the ramps increases with decreasing resolution. The benchmark method (blue) shows CAQE ramp errors that are generally higher across all resolutions and profiles because the averaging smooths out ramps. The linear interpolation method (grey) performs similarly to the benchmark, as it also smooths out ramps, flattening the profile. The median method (orange) displays similar CAQE ramp scores to the benchmark at the 2-hourly resolution across all profiles. The performance improves at 3 and 4-hourly resolution, particularly for the solar profile. The maximum (green) and minimum (red) methods consistently have higher CAQE ramps than the benchmark across most profiles and resolutions. The maximum method, for instance, has a lower CAQE for the ramps than the benchmark for all profiles except wind offshore. This method highlights ramp changes by using the extreme values within each aggregation block. However, this can also lead to significant differences in ramp distribution when they do not represent the typical behaviour of the profile. For example, the minimum method results in particularly higher CAQE ramp values in the solar profile, where the daily peaks are missed when the minimum value is used.

Similar to the CAQE for energy value distribution, the midpoint (pink), first (purple), and last (brown) methods have lower CAQE ramp scores than the benchmark across different profiles and resolutions. Specifically, the midpoint method in the solar profile exhibits low CAQE ramp scores, indicating a better representation of the ramp distribution. The first method selects the first value within each aggregation block and can capture the onset of a ramp or change within the profile. Likely because of this, the method has low CAQE ramp scores. Similarly, the last method captures the final value within each block and can capture the final direction of a ramp. This method also has relatively low CAQE ramp scores, especially in the wind onshore and offshore profiles at higher aggregation levels.

4.1.2. ***Hybrid***

Figure 8 illustrates the CAQE when the hybrid method is applied at different aggregation levels. The hybrid method integrates the benchmark averaging and maximum methods discussed in the previous section using different thresholds, aiming to achieve a better balance between accurately capturing the energy value and the ramp distributions of the original profile. The figure shows how using these different thresholds affects the CAQE for each profile. As the threshold increases, the benchmark is used more frequently (see Section 2.3), leading to lower CAQE values. In contrast, with lower thresholds the maximum method is used more often (see Section 2.3), resulting in higher CAQE values. This trend is consistent across all profiles and resolutions and is similar to the trends in CAQE observed for the benchmark and maximum methods in Figure 6. Except for the higher thresholds at the 2-hourly resolution and the 3 and 4-hourly resolutions for the wind profiles, the benchmark method has lower CAQE scores than the hybrid method. This suggests that, generally, the hybrid method introduces more errors in capturing the energy value distribution of the original energy demand profile, but at higher thresholds, it introduces lower errors in the energy value distribution in wind profiles.

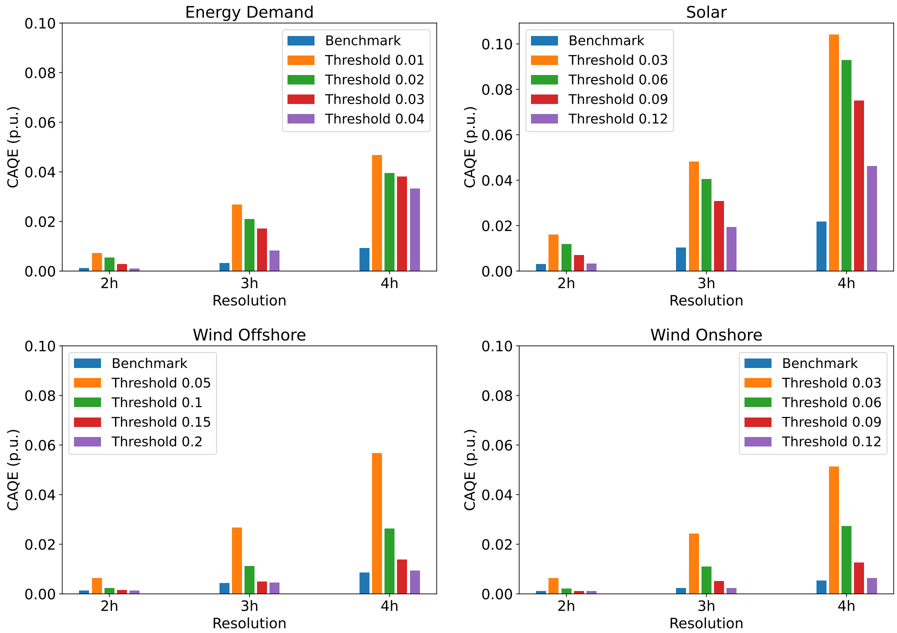


Figure 8: Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods using different thresholds, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

Figure 9 shows the CAQE of the ramps when the hybrid method is applied. Again a general trend can be observed where for all profiles with lower resolution higher errors in ramp representation occur. The trends observed for the energy demand and solar profiles are similar, to illustrate at 2-hourly resolution, an increase of CAQE ramp can be observed with higher thresholds. The lowest threshold for the energy demand profile shows a lower error than the benchmark (blue), while the first three thresholds of the solar profile have a lower error than the benchmark. This can be explained by the sensitivity of the hybrid method to the threshold. When using lower thresholds in the hybrid method, the maximum method is used more often to aggregate a block and the maximum is better at capturing extreme values and therefore better captures ramps. At 3-hourly resolution, this trend shifts slightly, as for both the energy demand and solar profiles, the CAQE ramp score first becomes lower with an increasing threshold reaching a minimum and finally increasing again. This may suggest that there is an optimal threshold that balances the use of the benchmark and the maximum. At the 4-hourly resolution, all thresholds have lower CAQE ramp scores than the benchmark. Interestingly, lower thresholds appear to result in lower ramp distribution errors. suggesting that upon lowering the resolution. This suggests that the maximum method better preserves ramps that might be underrepresented if the benchmark method was used.

For the wind offshore profile, the trend is different. At all resolutions, the ramp distribution errors are relatively stable. They become higher than the benchmark upon higher thresholds and drop again for the highest threshold. Only the last threshold shows a slightly lower error than the benchmark method. A similar trend can be observed for the wind onshore profile at the 2-hourly resolution where the CAQE ramp error remains similar with different thresholds, though slightly lower than the benchmark error. As the threshold increases, a slight rise in CAQE ramp scores can be observed for intermediate thresholds, followed by a decrease

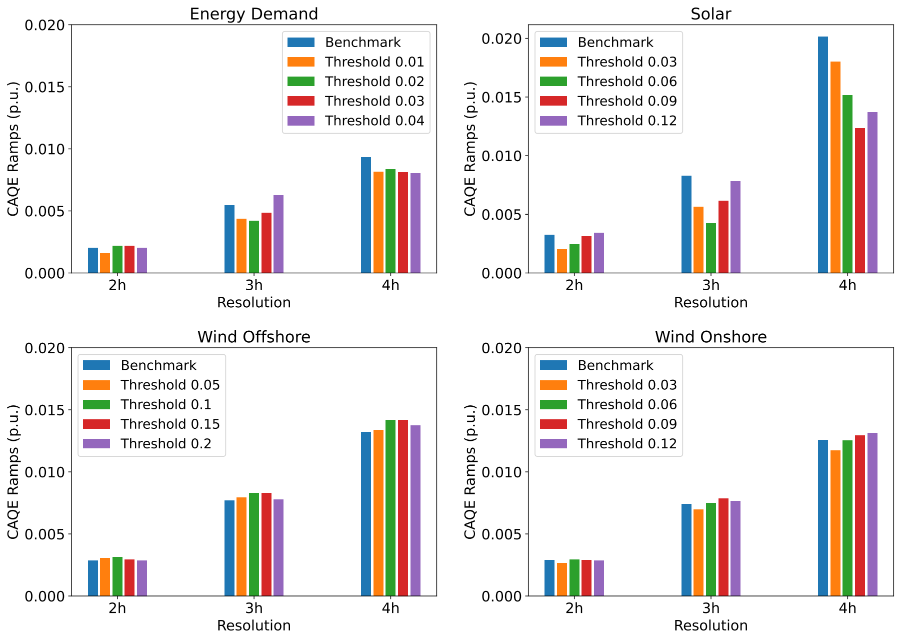


Figure 9: Cumulative absolute quantile error of the ramp distribution in the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods using different thresholds, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

for higher thresholds. This trend is consistent at the 3 and 4-hourly resolutions, where the lowest thresholds have the lowest ramp distribution errors. This difference between the wind onshore and offshore profiles is likely due to the slightly different ramp distribution as can be observed in Appendix J.40 and J.36. The wind onshore profile appears to have more frequent but moderate ramps, while the wind offshore profile has less frequent but larger ramps.

4.1.3. ***Conclusion Initial Representation Analysis***

The median, midpoint, first and last methods capture both the energy value and ramp distribution well, as indicated by the low CAQE and CAQE ramps. Moreover, generally, they outperformed the benchmark method showing lower differences with the original time series. In contrast, the maximum, minimum and linear interpolation methods showed higher errors with the original time series which suggests that they are less effective at representing the energy value and ramp distribution. Especially for the maximum and minimum methods substantially higher errors in CAQE and CAQE ramps were observed compared to the benchmark. For the hybrid method, the initial representation analysis pointed out that the errors in energy value and ramp distribution are dependent on the thresholds. For the representation of the energy value distribution, it proved to be difficult to match the original profiles, with generally significantly higher errors or similar errors to the benchmark method. For the ramp distribution representation in the energy demand and solar profiles, the hybrid method showed lower scores than the benchmark, particularly at the 2-hourly resolution. At 3 and 4-hourly, intermediate and higher thresholds often showed the lowest scores. In contrast, the wind offshore profile does not benefit significantly from the hybrid method across most thresholds, with the errors for both energy value and ramp distribution generally remaining close to or above the benchmark. The wind onshore profile shows better performance, particularly at lower thresholds.

4.2. ***Model Results***

In this section the impact of DS on model performance in the NWEU case study is assessed. The analysis begins with an evaluation of the basic statistics methods, followed by the hybrid method.

4.2.1. ***Basic Statistics***

The model results for the investment model configuration are summarised in a matrix in Figure 10. The x-axis of the matrix displays the evaluation metrics, including the total system costs or objective value, the solve time, as well as the investment decisions and flow and net position behaviour of the model. The y-axis represents the different DS experiments, where all input time series were down-sampled with the method specified. The y-axis is subdivided by the resolution the experiments were performed on. The values in the matrices represent the percentage differences compared to the hourly reference model, see Section 2.4.

It can be observed that at 2-hourly resolution for the objective value most of the methods have low percentage differences compared to the hourly reference, ranging from 0 to 14.9 % difference. Upon moving to lower resolutions the percentage difference with the hourly reference becomes higher and more methods start to perform worse than the benchmark indicated in grey. This indicates that for the objective value or total system costs the effects of the DS intensify with lower resolution. This is especially the case for the maximum and minimum methods, where on 2-hourly resolution the percentage difference is at 4.3 and 4.8 % compared to 0.5 % for the benchmark. On 3-hourly resolution, the difference rises to 7.7 and 9.2 % and at 4-hourly resolution to 13.2 and 14.9 % compared to 1.2 and 1.7 % for the benchmark, respectively. This trend in performance is in line with the initial representation analysis in Section 4.1.1, where the maximum and minimum exhibited high errors in energy value and specifically the minimum showed high ramp distribution errors for the solar profile. The linear interpolation and median have a much smaller percentage difference of both 0.5 % at 2-hourly resolution, performing similarly to the benchmark. At 3 and 4-hourly resolutions the median moves to outperforming the benchmark with a difference of 0.1 % and the linear interpolation moves to performing worse than the benchmark with a difference of 1.8 %. Again this is in line with the results of the analysis in Section 4.1.1. Here the linear interpolation method has similar scores to the benchmark in energy value and ramp distribution representation at 2-hourly resolution, but when moving to 3 and 4-hourly resolution the errors become higher than those of the benchmark. For the median, the errors become lower than that of the benchmark for both the energy value and ramp distribution representation at 3 and 4-hourly resolution.

Finally, the methods with the lowest difference compared to the hourly reference are the first, last and midpoint methods. At 2-hourly resolutions the difference is around 0.1 %, outperforming the benchmark. This trend persists at lower resolutions, however, at the 4-hourly resolution, it can be observed that the scores increase and the midpoint method performs at the same accuracy as the benchmark. This aligns closely with the analysis in Section 4.1.1, where these three methods consistently outperformed the benchmark across all profiles and resolutions in terms of energy value and ramp distribution errors. It must be noted that since the initial representation analysis in Section 4.1.1 was based solely on profiles from the Netherlands, slight differences in profiles from other countries may lead to (minor) variations in how DS methods affect energy value and ramp representation, and consequently, the model results. Despite these variations, the overall trends are consistent with the initial representation analysis.

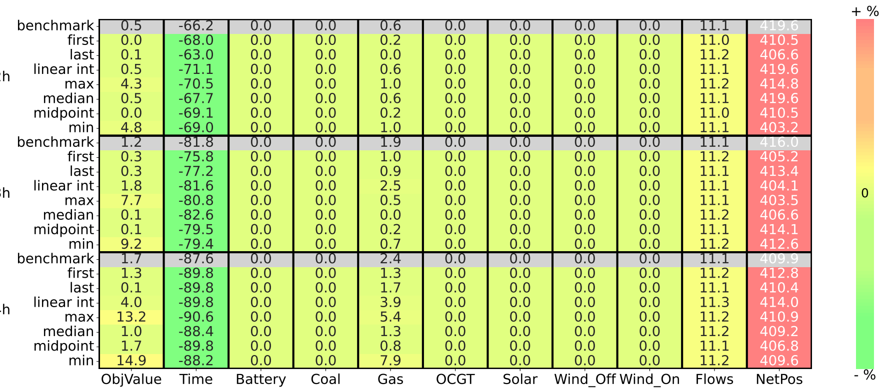


Figure 10: Overview of the results of the basic statistics down-sampling experiments in the investment model. On the x-axis the evaluation metrics are displayed including the objective value (total system costs), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. As opposed to the other metrics in the matrix the flow and net position behaviour indicated in this graph is the behaviour in the model after reallocation to 1 hourly level.

The solve time in all experiments and on all resolutions is significantly lower than 146.0s solve time in the hourly reference ranging from -90 to -66 % difference. This indicates that the solve time improves when using the down-sampled profiles in the model. It can also be observed that upon lowering the resolution the difference in solve time becomes larger. On 2-hourly resolution, the solve time drops by approximately 68 %, and on 3 and 4-hourly resolution it drops by approximately 80 and 88 %, respectively.

The differences in investment decisions show that DS has a relatively low impact. Generally, the experiments exhibit no difference with the hourly reference and for gas investments, the difference ranges from 0.2 to 7.9 %. Moreover, most methods either perform close to the benchmark or outperform it. The differences become more pronounced at lower resolutions, moving from 1 % to 2.5 and 7.9 % for the highest scores at 2, 3 and 4-hourly resolutions, respectively. The differences primarily occurring in the investment decisions for the gas asset is likely because the system compensates for the DS of RES profiles which reduces their availability. At the same time, the investment potentials for RES assets are fully met in this model configuration because of their low variable costs, see Figure L.45 in the Appendix for the investment decisions of the hourly reference model. Consequently, any generation mismatch caused by reduced RES availability needs to be accounted for, and the gas asset is favoured due to its lower investment costs compared to coal and its lower variable costs compared to OCGT.

For the different DS experiments, different effects come into play. In the case of the benchmark method, the averaging effectively represents the energy value distribution, as shown in the initial representation analysis. However, it performed worse in capturing the ramp distributions compared to other methods. This results in an underestimation of the required investments in gas assets compared to the hourly reference, as the less distinct ramps suggest fewer sudden changes that would require the flexibility typically provided by gas assets, see Figure L.46 in the Appendix for a comparison of the non-relative investment decisions. For the median and linear interpolation method, the root of this effect lies in the same nature as for the benchmark method, this is also why on 2-hourly resolution the performance of 0.6 % difference with the reference is very close to that of the benchmark. In the case of the maximum method, the energy value distribution is overestimated causing less need for extra capacity to handle the generation mismatch, leading to an underestimation of the gas investment decisions, as can be seen in the non-relative values in Figure L.46 in the Appendix. For the minimum method, the exact opposite is at play, here the ramps and energy values are underestimated. Especially on the 4-hourly resolution, the poor representation of energy values and ramp distributions of both the maximum and minimum is reflected by the higher differences of 5.4 and 7.9 %. The first, last and midpoint methods consistently show a low difference with the hourly reference and outperform the benchmark method in the investment decisions. The non-relative investment decisions in Figure L.46 in the Appendix, also show that the investment decisions of these methods are consistently very close to those of the hourly reference and this is likely because of the good representation of energy values and ramps as also came out from the initial representation analysis.

Overall, the performance of the methods in investment decisions and objective value largely aligns with the initial representation analysis results. However, one notable exception is that at the 3-hourly resolution, the maximum and minimum methods have differences of 0.5 and 0.7 % respectively and outperform the benchmark method, which is unexpected. Still, the same effect for the maximum and minimum methods occurs, they overestimate energy value leading to an underestimation of the gas investment decisions and underestimate energy value and ramps leading to an overestimation of gas investment decisions, respectively. The benchmark also underestimates the gas investment decision compared to the hourly reference but it appears to be that the difference here at 3-hourly resolution is higher than the difference occurring for the minimum and maximum methods. The 3-hourly resolution seems to provide a sweet spot where the effects of DS via the maximum and minimum methods align in a way that better supports investment decisions.

The behaviour of flows and net position was analysed after reallocation to hourly resolution, see Section 2.4.2. The results in Figure 10 show that the effect of DS on the flow behaviour is moderate and stays stable at around 11 % difference from the reference across different methods and different resolutions. In contrast, small differences in investment decisions and moderate differences in flows observed appear to lead to a significant difference in net position. The difference in net position is in the range of 403 to 419 % and appears to decrease upon lowering the resolution, where at 2-hourly resolution the highest difference is 419.6 % and at 3 and 4-hourly resolution the highest differences are 416 and 414 %. It must be noted, however, that on 2 and 3-hourly resolution most experiments outperform the benchmark.

Figure 11 shows the results of the experiments in the investment model with increased limits. The investment potential for solar, wind offshore, and wind onshore was lifted to get a better insight into the impact of DS on investment decisions, see Section 3. The results show that the DS now has a more noticeable impact as differences emerge not only in the gas assets but also in the wind offshore and coal assets. Still, the results of the increased limits investment model show a trend similar to the original investment model for both the objective value, solve time and investment decisions. Generally, the first, last, and midpoint methods show low differences with the reference and in all resolutions they consistently outperform the benchmark. The median method shows higher differences with the reference but still surpasses the benchmark at the 3- and 4-hourly resolutions, while the minimum and maximum methods show higher differences and do not outperform the benchmark. This aligns with the initial representation measurements. Notably, the percentage differences compared to the hourly reference become higher when moving to lower resolutions for the investment decisions and objective value. Again the solve time improves upon moving to lower resolutions and is significantly lower than the 120.3s solve time in the hourly reference ranging from -91 to -67 % difference.

The changes in investment decisions for wind offshore and coal are driven by the same factors as in the original investment model. Specifically, the system adjusts to changes in the representation of energy value and ramps in the down-sampled input profiles. The investment potentials for solar and wind onshore are already fully met because of their low variable costs, see Figure L.45 in the Appendix for the investment decisions of the hourly reference model. Consequently, any generation gaps need to be accounted for and

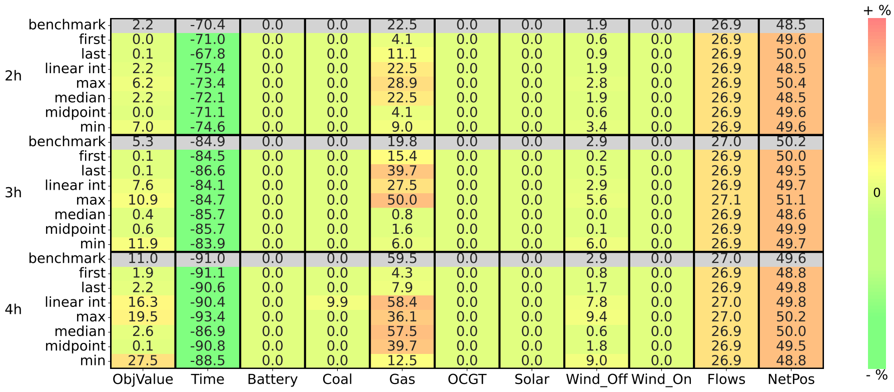


Figure 11: Overview of the results of the basic statistics down-sampling experiments in the investment model with increased limits. On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model, and the flow and net position behaviour in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value, investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2.

the mismatch is initially addressed with wind offshore investments, which likely contribute more to base load capacity considering its stable generation and low variable cost. However, the remaining mismatches, likely those related to peak demand are accounted for through adjustments in gas investments. This also explains why the differences between the down-sampled models and the hourly reference are more significant for gas investments. The difference in gas investments compared to the reference range from 4.1-59.5 %, compared to 0.1-9.4 % for wind offshore. Finally, only in the case of the linear interpolation method, additional adjustments are made to the coal investment decisions.

For the benchmark less investment is done in gas, and wind offshore likely because of the smoothing out of ramps as explained earlier resulting in an absolute difference of 22.5 %, see Figure L.47 in the Appendix for the non-relative investment decisions of the increased limits investment model. The same effect is at play for median and linear interpolation also reflected in the percentage differences with the reference that are the same as the benchmark at 22.5 % and 1.9 % for gas and wind offshore investment decisions, respectively. For the minimum and maximum method, the investment decisions are overestimated and underestimated respectively as explained previously which can also be observed in Figure L.47 in the Appendix. Overall, the trends in the investment decisions of wind offshore match the trend in objective value and the trends observed in the original investment model. Moreover, this trend can be linked again to the initial representation analysis where the first, last and midpoint exhibited lower energy value and ramp representation errors than the benchmark. On the 3 and 4-hourly resolution, this is also the case for the median method. The maximum and minimum consistently showed higher representation errors than the benchmark, especially for energy value representation and the linear interpolation method showed higher errors on the 3 and 4-hourly resolution. In contrast, for the gas investment decisions, the differences with the reference are higher as pointed out before, and especially at 4-hourly resolution, all experiments outperform the benchmark. On 2 hourly resolution, the same happens, although the linear interpolation method and median method perform very close to the benchmark method and the maximum shows a higher difference at 28.9 % compared to 22.5 %. On 3-hourly resolution, the last, linear interpolation and maximum methods show high differences with the reference ranging from 27.5 to 50 %. This is also shown in the investment decision differences in Figure L.47 in the Appendix. Moreover on the 4-hourly resolution minimum and maximum methods perform significantly better than previously, with differences of 36.1 and 12.5 % compared to 59.5 % for the benchmark. The sudden improvement in the performance of the maximum and minimum methods for gas investment decisions at the 4-hourly resolution may result from the role gas assets play in responding to peak demand. At this lower resolution, the maximum and minimum methods may better reflect ramps than the benchmark which becomes more beneficial considering this role of the gas asset. Still, the overall higher errors of most experiments indicate that at lower resolutions accurate investment decisions for the gas assets remain challenging.

The results in Figure 11 show that the effect of DS on the flow behaviour is more distinct than in the original investment model, but still stays stable at around 27 % difference from the reference across the different methods and different resolutions. The difference in net position is much less pronounced than in the original investment model varying from 48.5 to 50.4 % across all resolutions and methods. On the 3 and 4-hourly resolution, most experiments slightly outperform the benchmark.

Figure 12 presents the results of the operational model. The approximation of accuracy on hourly resolution is evaluated based on the objective value or total system costs, solve time, generation flows, and net position. The trend observed in the objective value aligns with the trend in the investment models. The first, last, and midpoint methods show low differences with the reference ranging from 0.3 to 2.1 % and consistently outperforming the benchmark across all resolutions. The median and linear interpolation methods perform similarly to the benchmark at the 2-hourly resolution, with a 1.8 % difference. However, as the resolution decreases to 3 and 4-hourly resolution, the median method begins to outperform the benchmark with a 1.2 % difference compared to the reference, while the linear interpolation method performs worse with 2.7 % compared to 2.3 % for the benchmark. The minimum and maximum methods exhibit high differences with the reference ranging from 3.3 to 14.1 % and consistently perform worse than the benchmark across all resolutions. Additionally, the percentage differences relative to the hourly reference model increase as the resolution decreases. The solve time improves upon moving to lower resolutions and is significantly lower than the 23.0s solve time in the hourly reference ranging from -86.4 to -60.8 % difference.

For the generation flows, the differences with the reference model range from 19 to 28 % and no single method consistently outperforms the benchmark across all resolutions. However, the difference in performance between the methods and the benchmark is relatively small, generally within a 1-5% range. At the 2-hourly resolution, the linear interpolation, first, median, and midpoint methods perform similarly to the benchmark. At the 3-hourly resolution, the median and midpoint methods with differences of 21.1 and 21.2 % slightly outperform the benchmark showing a 21.5 % difference. The linear interpolation method again performs close to the benchmark with 21.7 %. At the 4-hourly resolution, the median method with a 23.5 % difference slightly outperforms the benchmark with 23.6 %, while the midpoint and linear interpolation methods maintain performance around 27 % difference, close to the benchmark. The first and last methods, which previously performed well, show higher differences with the reference at the 4-hourly resolution, similar to the maximum and minimum methods. This decline is unexpected given the trends observed in the investment models and initial representation analysis. Overall, all DS experiments appear to have a similar effect on the generation flows. However, a breakdown of the flows by country and resolution shows the differences in behaviour for different DS methods. Generally, the differences in flows are due to the down-sampled RES profiles, this is shown by these flows being lower in size than in the hourly reference model, see Figure K.41 in the Appendix. In turn, this reduction in RES flows leads to increased reliance on balancing assets such as batteries, OCGT, and gas, see K.41 in the Appendix. This trend is consistent across all DS methods but is more pronounced in methods like maximum and minimum compared to the benchmark, see K.42 in the Appendix.

The differences with the reference are high for the net position, ranging from 11.2 to 116.2 %. Strikingly, at 3-hourly resolution, the differences are much lower than at 2 and 4-hourly resolution ranging between 11.2 and 34.5 %, compared to ranging between 76.7-116.2 %. At 2-hourly resolution, the last and minimum methods with differences of 76.7 and 80.9 % outperform the benchmark that has a difference of 84.1 %, while the median performs similarly at 84.1 %, and other methods like the first, linear interpolation, and midpoint are slightly worse with 84.3, 84.2 and 84.3 %. At the 3-hourly resolution, the last method again outperforms the benchmark with a difference of 11.2 % compared to 13.6 %, respectively. The maximum method has a difference of 34.5 % compared to the 13.6 % difference of the benchmark and the other methods perform close to the benchmark

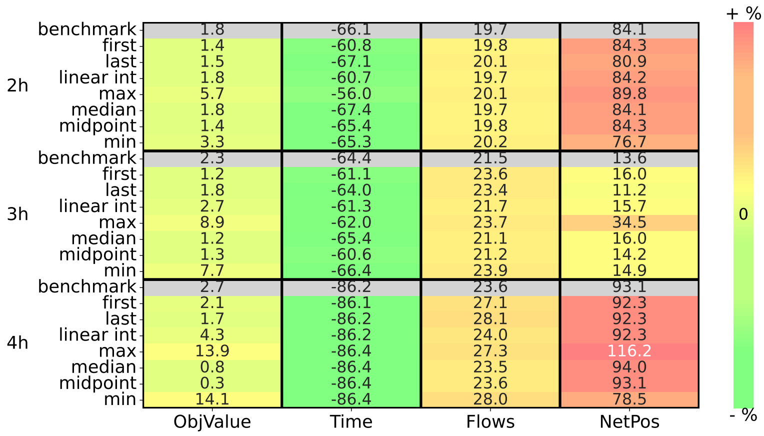


Figure 12: Overview of the results of the basic statistics down-sampling experiments in the operational model. On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the flows and net position. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not.

with differences ranging from 14 to 16 %. At the 4-hourly resolution, methods like the first, last, linear interpolation, midpoint, and minimum all outperform the benchmark, with the minimum method showing the lowest percentage difference of 78.5 % compared to the benchmark at 93.1 %. The maximum method has a significantly higher difference of 116.2 %. A similar trend is observed in the increased limits investment model where on 4-hourly resolution most methods outperform the benchmark. Likely the same effect, of smaller differences in flows leading to much higher differences in net position, here is at play. The results show that in the net position, small errors in generation due to DS can lead to significant shifts in import and export decisions, as these are influenced by relative differences between regions. Methods that capture critical points within an aggregation block, such as the last first and midpoint methods, may align better with the actual conditions affecting import and export flows, especially at the end of a time block. The minimum method, by focusing on the lowest values, may better capture periods of minimum availability, leading to more conservative export and import decisions that align more closely with the hourly reference. This analysis suggests that while the impact of DS is relatively minor on intra-country dynamics (flows), it becomes more significant at the inter-country dynamics (net position). The operational model struggles more with matching the hourly generation of the reference model in terms of flows and net position. While it struggles more with outperforming the benchmark in terms of flows and intra-country dynamics, at higher resolutions, it is easier to outperform the benchmark in net position and inter-country dynamics.

4.2.2. ***Hybrid***

Figure 13 presents the results of the hybrid experiments in the investment model. The x-axis shows the different thresholds used in the hybrid method and the model metrics that are evaluated. The y-axis shows the different profiles that were down-sampled and it is further subdivided in the resolutions used for DS. Please refer to Appendix G for the thresholds used for each profile on different resolutions. For the objective value or total system cost the percentage differences with the reference range from 0.5 to 22 %. With decreasing resolution, the difference becomes larger, which is most pronounced for the experiments with the energy demand and solar profiles. The benchmark method generally performs better than the hybrid method in terms of the objective value with percentage differences ranging from 0.5 to 1.7 % on all resolutions. Only the higher thresholds in the experiments with wind profiles perform close to the benchmark on all resolutions. On the 2-hourly resolution only the highest threshold for experiments with the solar profile and on 4-hourly resolution the highest threshold in the experiments with the energy demand profile performed with similar differences compared to the reference as the benchmark. Overall, a decrease in percentage difference occurs with larger thresholds. This suggests that increasing the threshold improves the accuracy approximation of the hourly reference in terms of objective value. This aligns with the initial representation analysis, which showed that higher thresholds achieve a better balance between energy values and ramp distribution representation.

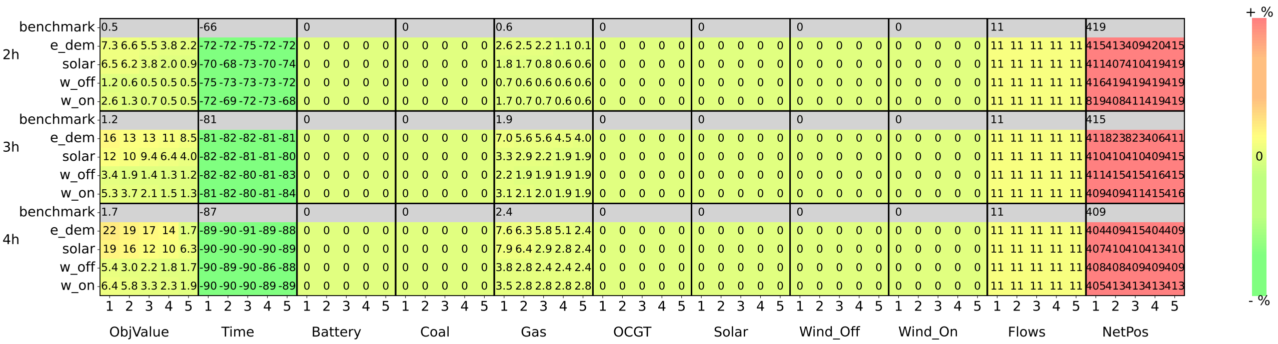


Figure 13: Overview of the results of the hybrid down-sampling experiments in the investment model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2.

The solve time is significantly lower than the 146.0s in the hourly reference ranging from -90 to -66 % difference, indicating that the solve time improves when using the down-sampled profiles in the model. It can also be observed that upon lowering the resolution the difference in solve time becomes larger. On 2-hourly resolution, the solve time drops by approximately 70 %, and on 3 and 4-hourly resolution it drops by approximately 80 and 90 %, respectively. The drop in solve time in the hybrid experiments is similar to that of the benchmark, which is 66 % on 2-hourly resolution and 81 and 87 % on 3 and 4-hourly resolution.

In terms of investment decisions, the differences arising from DS are minimal. The hybrid method does not significantly change investment decisions compared to the hourly reference, likely because all renewable potentials are met. The generation mismatch that arises because of DS is primarily addressed by adjusting investments in gas as can be observed from the percentage differences ranging from 0.6 to 7 %. Here, the hybrid method shows promise in outperforming the benchmark, particularly at the highest threshold for the energy demand profile on 2-hourly resolution with a difference of 0.1 % compared to a 0.6 % difference for the benchmark. The highest thresholds for solar and wind profiles perform at the same accuracy as the benchmark. At 3-hourly resolution the benchmark is not outperformed, however, it must be noted that for offshore wind almost all thresholds perform similarly to the benchmark at 1.9 %, as well as the higher thresholds for the solar and wind onshore profiles. At the 4-hourly resolution, the highest thresholds for solar, energy demand and wind offshore profiles perform at the same accuracy as the benchmark at 2.4 % difference and the highest thresholds for wind onshore, perform with slightly higher differences to the benchmark of 2.8 %. Figure L.48 in the Appendix shows that experiments with the energy demand profile tend to overestimate gas investments, while solar profile experiments tend to underestimate them. The wind offshore and onshore profiles exhibit relatively stable performance with slight underestimation in gas investments.

When linking these trends in investment decisions to the initial representation analysis, it shows that thresholds providing the best balance between the lowest possible error in energy value and ramp distribution representation generally yield the lowest differences with the hourly reference. For example, higher thresholds for energy demand, solar, and wind onshore profiles at the 2-hourly resolution show low differences in investment decisions. Specifically, the highest threshold for energy demand results in the lowest CAQE error and a CAQE ramp error slightly lower than the benchmark, see Figures 8 and 9. Similarly, for the solar profile, the fourth threshold performs well, and for wind onshore, the highest thresholds show strong results. At the 3-hourly resolution, however, it becomes more challenging to outperform the benchmark. This trend is somewhat expected for the wind onshore and wind offshore methods, where CAQE ramp errors at this resolution are similar to or higher than the benchmark. The same applies to energy value distribution errors. Although there is a significant improvement in ramp error for the energy demand and solar profiles at 3-hourly resolution for intermediate thresholds, this improvement appears to be insufficient to offset the higher energy values errors associated with these thresholds. Leading to a higher difference with the hourly reference model performance. Interestingly, at the 4-hourly resolution, methods that are better at representing ramps tend to outperform the benchmark. Possibly in such cases, better ramp representation may become more critical, allowing these methods to surpass the benchmark as general trends that are well-maintained through averaging are already smoothed out on lower resolution.

The behaviour of flows and net position was analysed after reallocation to hourly resolution, see Section 2.4.2. The results in Figure 13 show that the effect of DS on the flow behaviour is moderate and stays stable at around 11 % difference. In contrast, small differences in investment decisions and moderate differences in flows lead to a significant difference in net position. The difference in net position is in the range of 404 to 823 %. Specifically, the lowest threshold for wind onshore reaches a difference of 819 % with the reference, similar to the second and third threshold for the energy demand profiles on 3-hourly resolution with a difference of 823 %, in both cases this is much higher than the benchmark. It must be noted, however, that on 2 and 3-hourly resolution most experiments outperform the benchmark or perform at the same accuracy as the benchmark.

Overall, no single method consistently outperforms the benchmark across all evaluation metrics. However, if the primary focus is on matching investment decisions with those of the hourly reference model, the higher thresholds for the energy demand, solar, and wind onshore profiles perform well at the 2-hourly resolution. On 3-hourly resolution, the highest thresholds for the solar and wind profiles perform well. Similarly, at the 4-hourly resolution, higher thresholds for the energy demand, solar and offshore wind profiles also show strong performance.

Figure 14 presents the results of the investment model with increased limits. Most investment decisions remain the same as in the hourly reference for battery, OCGT, solar and wind offshore. However, now the differences have emerged not only in the gas assets as observed in Figure 13 but also in the wind offshore and coal assets. The trend in performance has also become different in the objective value. DS the energy demand profile is performing significantly better over all three resolutions, with on 2-hourly resolution the two highest thresholds outperforming the benchmark with differences of 1.1 and 0.3 %. On 3-hourly resolution, the two highest thresholds outperform the benchmark as well, with differences of 4.2 and 1.8 % and on 4-hourly resolution all thresholds expect the highest outperform the benchmark method. The wind offshore experiments perform the same as the benchmark for all resolutions. For the wind onshore profiles, a similar trend can be observed for the higher thresholds performing the same as the benchmark. The solar profile exhibits the largest differences with the hourly reference, on 2-hourly resolution they range from 2.5 to 7.5 %, on 3 and 4-hourly resolution they range from 6.1 to 13 % and 12 to 23 %, respectively. Overall, the percentage difference decreases

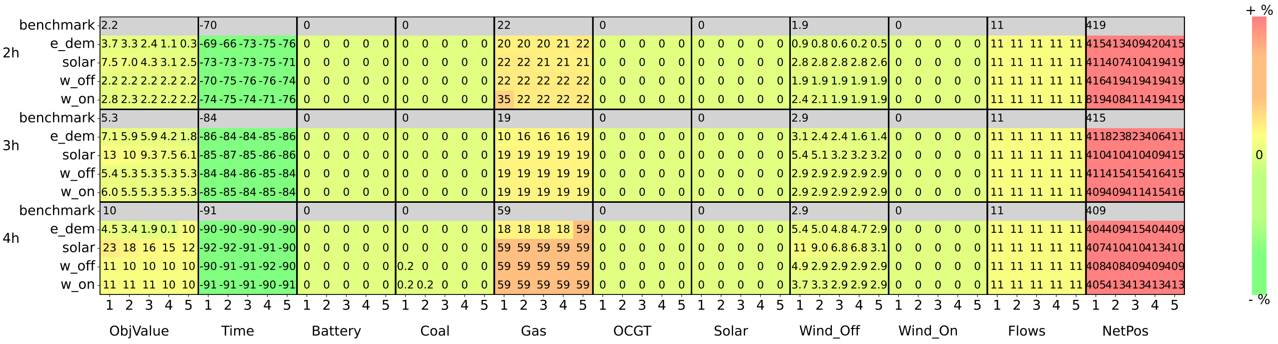


Figure 14: Overview of the results of the hybrid down-sampling experiments in the increased limits investment model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value (total system cost), investment decisions, flows and net position the differences are absolute while for the solve time, they are not. As opposed to the other metrics in the matrix the flow and net position are the behaviour in the model after reallocation to 1 hourly level, see Section 2.4.2.

upon using higher thresholds, which again suggests that increasing the threshold improves accuracy in terms of objective value. Again the solve time improves upon moving to lower resolutions and is significantly lower than the 120.3s solve time in the hourly reference ranging from -91 to -67 % difference.

The changes in investment decisions for gas, wind offshore and coal are driven by the same factors as in the original investment model. Specifically, the system adjusts for mismatches in generation and demand that are introduced by the DS of the input profiles. The investment potentials for solar and wind offshore are fully met, leading to the adjustments being made in wind offshore, gas and coal, see Figure L.45 in the Appendix for the investment decisions of the hourly reference model. The mismatch is initially addressed with wind offshore (and coal) investments, which likely contribute more to base load capacity considering its low variable cost. For coal the cost is lower in comparison with other fossil fuel generation assets. The remaining mismatches are managed through adjustments in gas investments. This also explains why the differences with the hourly reference are more significant for gas investments ranging from 10 to 59 %, compared to those of coal and wind offshore ranging from 0.2 to 11 %.

For the investment decisions in offshore wind, the energy demand profile with all thresholds outperforms the benchmark on 2-hourly resolution with differences between 0.2 and 0.9 % compared to a difference of 1.9 % for the benchmark. All thresholds for wind offshore and higher thresholds for the wind onshore profile have the same difference with the reference of 1.9 % as the benchmark. For the wind onshore profile, the higher thresholds perform at the same accuracy as the benchmark. The thresholds for the solar profile show higher differences with the reference than the benchmark ranging between 2.6 and 2.8 %. A similar trend persists on the 3-hourly resolution, where the highest thresholds outperform the benchmark for the energy demand profile and all thresholds for the wind profiles show the same difference with the reference as the benchmark at 2.9 %. The experiments with the solar profile have higher differences than the benchmark ranging from 3.2 to 5.4 %. On the 4-hourly resolution this completely changes, now the hybrid method generally performs at lower accuracy than the benchmark, except for the highest threshold for the energy demand profile and the higher thresholds for the wind offshore and wind onshore profiles that show the same difference of 2.9 % as the benchmark. Overall, with decreasing resolution, it becomes harder to beat the benchmark. Figure L.48 in the Appendix shows that the benchmark underestimates the investment decisions in wind offshore investment decisions, while the energy demand profile tends to overestimate gas investments. The solar profile experiments tend to underestimate the investment decisions for wind offshore substantially, while the wind profile experiments perform close to the benchmark and slightly underestimate the investment decisions.

For the gas investment decisions, the experiments either outperform the benchmark or show the same difference with the reference as the benchmark. On all resolutions, the first four thresholds used in the energy demand profile experiments have a lower difference with the reference than the benchmark. The improvement in the performance may be due to the role gas assets play in responding to peak demand. The initial representation analysis showed that the hybrid methods, although most at the cost of a lower accuracy in energy values representation, better reflect ramps than the benchmark. Because of the role of the gas asset in addressing peak demand a better representation of ramps may become more important than a good representation of the energy values, leading to the overall better accuracy for the gas investment decisions.

Similar to the original investment model the effect of DS on the flow behaviour is moderate and stays stable at around 11 % difference from the reference. The difference in net position is in the range of 404 to 823 %. Specifically, the lowest threshold for wind onshore reaches a difference of 819 % with the reference, similar to the second and third threshold for the energy demand profiles on 3-hourly resolution with a difference of 823 %. It must be noted, however, that on 2 and 3-hourly resolution most experiments outperform the benchmark or perform at the same accuracy as the benchmark.

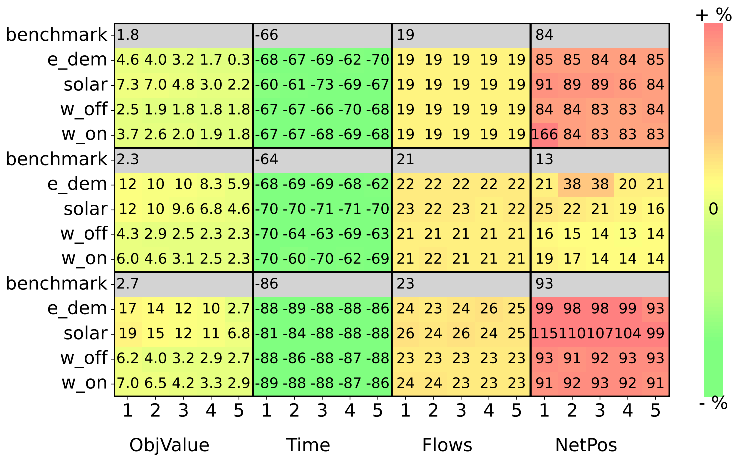


Figure 15: Overview of the results of the hybrid DS experiments in the operational model. On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The thresholds are ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value (total system cost), solve time, flows and net position. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not.

Figure 15 presents the results of the operational model. The trend observed in the objective value or total system cost aligns closely with that seen in the investment models. On the 2-hourly resolution, the highest thresholds for the energy demand profile outperform the benchmark with differences of 0.3 and 1.7 % compared to 1.8 %. Moreover, the highest thresholds for wind offshore and wind onshore perform close to the benchmark, with differences of 1.8 %. Again the difference with the reference of the solar profile experience is higher ranging from 2.2 to 7.3 %. On the 3-hourly resolution, the energy demand profile performs worse with differences ranging from 5.9 to 12 % compared to 2.3 % of the benchmark, but the performances of the wind offshore and onshore profiles remain close to the benchmark with differences of 2.3 %. On the 4-hourly resolution, the highest threshold for the energy demand profile shows the same difference with the reference as the benchmark as well as the highest thresholds for wind offshore and wind onshore. An increase in accuracy for each profile can be observed when using lower thresholds. The solve time improves upon moving to lower resolutions and is significantly lower than 23.0s solve time in the hourly reference ranging from -89 to -61 % difference.

For the flows, the percentage differences compared to the hourly reference are higher than for the objective value but remain more stable across the different resolutions. On the 2-hourly resolution, all methods perform very close to the benchmark which has a percentage difference of 19%. When moving to 3-hourly resolution, again the scores of all experiments are very close to that of the benchmark at 21 % ranging from 21 to 23 %. All thresholds of the offshore wind profile and the highest thresholds for the onshore wind profile show the same difference with the reference as the benchmark. Again for the 4-hourly resolution, the differences range from 23 to 26 % and are all very close to that of the benchmark at 23 %. All thresholds of the offshore wind profile and the higher threshold for the onshore wind profile show the same differences as the benchmark. In contrast to the trend for the objective value higher thresholds do not necessarily lead to lower differences with the reference, suggesting that preserving the energy values may become of less importance when attempting to match the flows of the hourly reference model. While the scores for the flows appear relatively stable for different profiles and different resolutions, a breakdown of the flows by country and resolution shows the differences in behaviour for different DS methods. Generally, the differences in flows are due to the down-sampled RES profiles, this is shown by these flows being lower than in the hourly reference model, see Figure K.44 and K.43 in the Appendix. For example for the solar experiments, it can be observed that the flow generation of solar becomes lower compared to the hourly reference, see Figure K.44 in the Appendix. This is consequently picked up by wind onshore and wind offshore. When DS the energy demand, the flow generation of all renewables becomes lower, see Figure K.43 in the Appendix.

Moving to the net position, a similar trend can be observed as for the flows, where the wind onshore and wind offshore experiments show similar or lower differences than the benchmark on all resolutions. The differences in net position range from 83 to 166 % on 2-hourly resolution with the lowest threshold for wind offshore showing the highest difference with the reference. On 3-hourly resolutions the differences range from 13 to 38 % and on 4-hourly resolutions the differences range from, 91 to 115 %, with the experiments of the solar profile exhibiting the largest differences. The higher differences at the 2 and 4-hourly resolutions suggest that it is easier to match the net position on the 3-hourly resolution, though it is easier to outperform the benchmark on the other two resolutions. The difference in behaviour between the flows and net position is likely because the flows are impacted less by potentially misrepresenting available energy, as these effects can be smoothed out by the operational flexibility of the system, resulting in smaller deviations from the hourly reference model. For net position, however, similar to what was observed for the basic statistics methods, small errors in generation due to the DS lead to significant shifts in import and export decisions. Again this analysis suggests that while the impact of DS is relatively minor on intra-country dynamics (flows), it becomes more significant at the inter-country dynamics (net position).

4.2.3. ***Conclusion Model Results***

Basic Statistics Investment

The analysis of the experiments in the investment model revealed that the basic statistics methods showed low differences with that of the reference for the objective value ranging from 0-14.9 %. When moving to lower resolutions, the difference with the reference becomes higher, at 2-hourly resolution the differences range from 0-4.8 %, on 3-hourly resolution from 0.1-9.2 % and on 4-hourly resolution from 0.1-14.9 %. Methods like the first, last and midpoint generally outperform the benchmark in terms of difference with the reference, with 0-0.5 % difference compared to 0.5 % for the benchmark at 2-hourly resolution. On 3 and 4-hourly resolutions their differences with the reference range from 0.1-0.3 % and 0.1-1.7 % compared to 1.2 and 1.7 % for the benchmark, respectively. Methods consistently exhibiting high differences are the minimum and maximum methods, with differences ranging from 4.3-4.8 %, 7.7-9.2 % and 13.2-14.9 % on 2, 3 and 4-hourly resolution respectively. The median and linear interpolation methods generally showed similar or slightly higher differences to the reference compared to the benchmark. This also holds for the differences in investment decisions observed, especially in the gas investments where the difference ranged from 0-1 %, 0.2-1.9 % and 0.3-7.9 % on 2, 3 and 4-hourly resolution. Again the first, last and midpoint methods showed lower differences than the benchmark, with values ranging from 0-0.6 %, 0-1 % and 0.8-1.7 % compared to 0.6, 1.9 and 2.4 % for the benchmark on 2, 3 and 4-hourly resolution, respectively. This trend in the difference between the reference and performance compared to the benchmark aligns with the results of the initial representation analysis where the first, last and midpoint showed lower energy value and ramp representation errors compared to the benchmark. In contrast the maximum and minimum showed significantly higher errors, especially for the energy value representation and the median and linear interpolation showed errors similar to those of the benchmark. The flow behaviour of all experiments showed to remain stable at around 11 % difference compared to the reference. The small differences observed in the investment decisions and moderate differences in the flows lead to a large difference in net position ranging from 403-419 %. Still at 2, 3 and 4-hourly generally the experiments outperformed the benchmark with a difference lower than 419.6 % ranging from 403.2-410.5 % at 2-hourly resolution, and 403.6-413.4 % compared to 416 % of the benchmark on 3-hourly resolution and 409.8-409.2 % compared to 409.4 % on 4-hourly resolution.

In the results of the investment model with increased limits similar trends were observed in the objective value where the differences ranged from 0-27.5 % and the first, last and midpoint method consistently outperformed the benchmark method with differences ranging from 0-0.1 % compared to 2.2 %, 0.1-0.6 % compared to 5.3 % and 0.1-2.2 % compared to 11.0 % on 2, 3 and 4-hourly resolutions. The maximum and minimum methods exhibited higher differences of 6.2 & 7 %, 10.9 & 11.9 % and 19.5 & 27.5 % on 2, 3 and 4-hourly resolution. Low differences were observed for the experiments in the wind offshore investment decisions ranging from, 0.1-9.4 % across all resolutions. A similar trend was observed where the first, last and midpoint methods outperformed the benchmark method with scores lower than that of the benchmark, 0.6-0.9 % vs 1.9 %, 0.0-0.5 % vs 2.9 % and 0.6-1.8 % vs 2.9 %. The difference in gas investment decisions is substantially higher ranging from 0.8-57.5 %. For gas investment decisions, it is harder to identify one method that consistently outperforms the benchmark over different resolutions. However interestingly, on 4-hourly resolution minimum and maximum methods perform significantly better, with differences of 36.1 and 12.5 % compared to 59.5 % for the benchmark. The sudden improvement likely results from the role gas assets play in responding to peak demand. At this lower resolution, the maximum and minimum methods may better reflect ramps than the benchmark which becomes more beneficial considering the role of the gas asset when responding to peak demand. Still, the overall higher errors of most experiments indicate that at lower resolutions accurate investment decisions for the gas assets remain challenging. Similarly to the original investment model, the flow behaviour remains stable across the different experiments varying around 27 % difference on all resolutions and the net position difference is higher around 50 % difference on all resolutions. Notably, the difference in net position is much lower in the increased limits investment model.

Basic Statistics Operation

The analysis of the experiments in the operational model revealed that the basic statistics methods showed low differences with the reference for the objective value ranging from 0.3-14.1 %. The differences become higher when lowering the resolution. On 2-hourly resolutions, they range from 1.4-5.7 %, on 3-hourly resolution from 1.2-8.9 % and on 4-hourly resolution from 0.3-14.1 %. The first, last and midpoint methods generally outperform the benchmark, with a 1.4-1.5 % difference compared to 1.8 % for the benchmark at 2-hourly resolution. On 3 and 4-hourly resolutions their differences with the reference range from 1.2-1.8 % and 0.3-1.7 % compared to 2.3 and 2.7 % for the benchmark, respectively. Again methods consistently exhibiting high differences are the minimum and maximum methods, ranging from 3.3 & 5.7 %, 7.7 & 8.9 % and 13.2 & 14.9 % on 2, 3 and 4-hourly resolution respectively. The median and linear interpolation methods generally showed similar or slightly higher differences to the reference compared to the benchmark. This trend aligns with the results of the initial representation analysis where the first, last and midpoint showed lower energy value and ramp representation errors compared to the benchmark, the maximum and minimum showed significantly higher errors, especially for the energy value representation. The median and linear interpolation showed errors similar to those of the benchmark.

For the flows, the differences with the reference are higher ranging from 18-28 % across different resolutions and increase upon lowering the resolution. No single method consistently outperforms the benchmark, with performance generally close to the benchmark. The differences with the reference for the net position are significantly higher, with the difference ranging from 76.7-89.8 %, 11.2-34.5 % and 78.5-116.2 % on 2, 3 and 4-hourly resolution respectively. On 2-hourly resolution, the last and minimum methods have a difference of 80.9 and 76.7 % compared to the 84.1 % of the benchmark. At 3-hourly resolution, the last method is the only one that outperforms the benchmark with 11.2 % compared to 13.6 % and at the 4-hourly resolution the first, last, linear interpolation and minimum also outperform the benchmark with differences ranging from 78.5-92.3 % compared to 93.1 %. The other methods show similar differences to the benchmark and the maximum method consistently shows higher differences on all resolutions specifically 116.2 % on 4-hourly resolution which is the highest difference. This trend suggests that the first, last, linear interpolation and minimum methods better capture critical moments in cross-border trade decisions, especially at lower resolutions, where averaging by the benchmark method may smooth out important fluctuations.

The operational model shows that while it is challenging to outperform the benchmark accuracy in terms of generation flows, especially at lower resolutions, the ability to manage net position improves with methods that focus on specific points or extremes. Moreover, for the operational model, the first, last, and midpoint methods show consistent performance in managing objective value and net position.

Hybrid Investment

The analysis of the investment models, showed that with the hybrid method, it is difficult to consistently outperform the benchmark accuracy for the objective value. In the original investment model, the difference with the benchmark ranges from 0.5-22 %, and increases upon lowering the resolution. In the increased limits investment model, the differences are in a similar range of 0.1-23 %. In the original investment model, the benchmark is rarely outperformed, with higher thresholds of the wind offshore and onshore profiles showing similar differences compared to the reference as the benchmark at 0.5, 1.2 and 1.7 %, at 2, 3 and 4-hourly resolutions. Only the highest threshold of the energy demand matches the benchmark difference at 4-hourly resolution with a 1.7 % difference. In the increased limits investment model, DS the energy demand profile significantly improves performance across all resolutions, where the two highest thresholds on 2 and 3-hourly resolution have differences of 0.3 and 1.1 % vs 2.2 % benchmark difference and 4.2 and 1.8 % vs 5.3 % of the benchmark, respectively. On 4-hourly resolution, all thresholds except the last outperform the benchmark at 10% difference, with differences ranging from 0.1-4.5 %. On all resolutions, the wind profiles also show differences close to the benchmark, especially at higher thresholds. Across both models, accuracy generally increases with higher thresholds, reflecting a better balance between energy value representation and ramp errors.

In the original investment model the differences in gas investments range from 0.1-7.9 % and only the highest threshold of the energy demand profile experiment on 2-hourly resolution shows potential to outperform the benchmark with a difference of 0.1 % compared to 0.6 %. On the other resolutions mainly the wind profiles perform stably close or at the same difference compared to the benchmark. The hybrid method shows greater promise in the increased limits investment model. Here, the hybrid method outperforms the benchmark accuracy for the wind offshore investments at 2 and 3-hourly resolutions, especially with the energy demand profile exhibiting differences between 0.5-0.9 % vs 1.9 % of the benchmark and 1.4-2.4 % vs 2.8 % of the benchmark, respectively. The solar profiles consistently under-perform with differences higher than the benchmark, while wind onshore and offshore profiles generally align with the benchmark. For gas investment, the first four thresholds exhibit lower differences than the benchmark on all resolutions ranging from 20-21 % vs 22 % of the benchmark and 10-16 % vs 19 % and 18 % vs 59 % of the benchmark on 2, 3 and 4-hourly resolutions respectively. Also, the solar profile's higher thresholds show lower differences than the benchmark at 21 % on 2-hourly resolutions and at all other resolutions the differences are the same as the benchmark. The wind profiles also show similar differences to the benchmark on all resolutions. Both investment models show a sudden improvement in the hybrid method performance at the 4-hourly resolution. This is likely due to better ramp representation in the hybrid method compared to the benchmark at lower resolutions, which appears to become critical for gas assets addressing peak demand, allowing it to surpass the benchmark accuracy.

In both models, the flow behaviour of all experiments showed to remain stable at around 11 % difference compared to the reference. The small differences observed in the investment decisions and moderate differences in the flows lead to a large difference in net position ranging from 403-823 %. It must be noted, however, that on 2 and 3-hourly resolution most experiments outperform the benchmark or perform at the same accuracy as the benchmark. In conclusion, the energy demand profile with intermediate to high thresholds shows the most promise with low differences compared to the reference and also lower differences than the benchmark. This is particularly the case in the increased limits investment model. When matching investment decisions only, higher thresholds for the energy demand and solar profiles show the most potential in the investment model. In the increased limits investment model, the energy demand profile experiments outperform the benchmark for both gas and wind investments. The wind onshore profile outperforms the benchmark in gas investments and matches the benchmark accuracy in wind investments and for the wind offshore profile the other way around holds.

Hybrid Operation

The analysis of the operational model, showed that in the hybrid experiments, the differences in objective value ranged from 0.3-19 %, with increasing differences upon lowering the resolution. With the hybrid method, it proved to be difficult to outperform the benchmark accuracy for the objective value. Only the highest thresholds for the energy demand profile outperform the benchmark at the 2-hourly resolution with 0.3 % difference compared to 1.8 %. While the highest thresholds for wind offshore and onshore perform close to the benchmark with differences of 1.8, 2.3 and 2.7 %. The differences are generally lower with higher thresholds for each profile.

The differences in generation flows range from 19-26 % across all resolutions for all experiments. At 2-hourly resolution, all experiments exhibit the same difference as the benchmark at 19 %. At 3-hourly resolution, all thresholds for the wind profiles exhibit the same difference as the benchmark at 21 %, and the solar and energy demand profiles experiments show higher scores ranging from 21-23 % difference. At 4-hourly resolution again the wind profiles experiments perform close to the benchmark at 23 % and the energy demand and solar experiments perform slightly higher with differences ranging from 23-26 %. Unlike the objective value, higher thresholds do not necessarily improve accuracy, suggesting that energy value representation may be less crucial for matching generation flows.

For net position, the differences observed for all experiments are higher ranging from 13-166 %. Only wind onshore and offshore profile experiments outperform the benchmark at 2- and 4-hourly resolutions with differences at 83 % vs 84 % of the benchmark and 91-92 % vs 93 % of the benchmark. The significantly higher percentage differences at 2- and 4-hourly resolution of 83-166 % compared to the lower ranges 13-38 % at 3-hourly resolution suggest that DS makes it easier to match the net position at 3-hourly, though outperforming the benchmark is easier at the other two resolutions. The higher percentage differences observed for the net position compared to flows indicate that the DS has smaller differences on intra-country level compared to higher differences in inter-country dynamics, particularly at 2- and 4-hourly resolutions. Overall, there is not one profile and threshold that shows consistently lower differences compared to the hourly reference than the benchmark across. However, when matching flows and net position is considered, the wind onshore and offshore profiles consistently outperform the benchmark at 2- and 4-hourly resolutions. These profiles also show differences compared to the reference that are close to the benchmark for objective value.

Solve Time

For all DS experiments and in all model configurations a decrease in solve time was observed. On 2-hourly resolution, the solve time dropped by approximately 60-70% and on 3-hourly resolution it dropped by 80-85 %. Lastly, on 4-hourly resolution, it dropped by approximately 86-90 % resolution. This highlights the potential for achieving faster solving times by lowering temporal resolution but at the same time approximating the accuracy of evaluation metrics on hourly resolution and even improving compared to the benchmark.

4.3. Quality Check

This section presents the results of the quality check, each down-sampled model was reallocated to hourly resolution and the energy not served (ENS) was evaluated.

4.3.1. Investment Models

Figure 16 shows the ENS results of the investment models using the basic statistics methods. The ENS is compared to that in the hourly reference model, which is at 815102 MWh and 203859 MWh, for the original and increased limits investment model respectively. For both the investment (left) and increased limits investment model (right), the ENS increases with lower resolution.

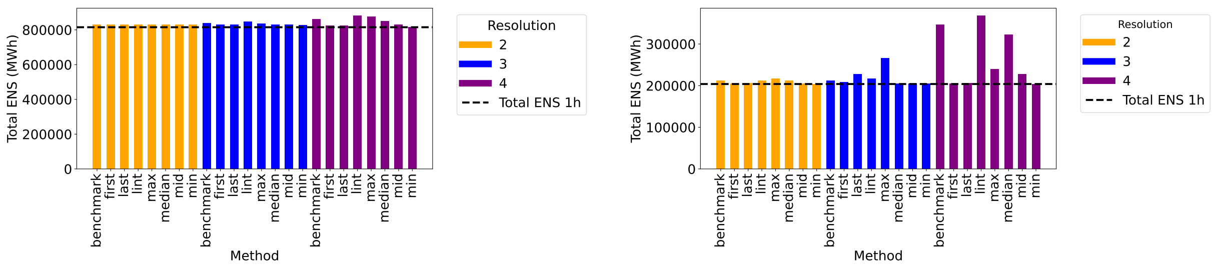


Figure 16: An overview of the energy not served after reallocation of the investment models (left) and investment model with increased limits (right) using the basic statistics methods. The energy not served is compared to that in the hourly reference model. The methods used for down-sampling are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions.

At the 2-hourly resolution, the ENS remains close to the reference model ranging around 815102 MWh and 203859 MWh in the original and increased limits investment model configurations, respectively. However, in the increased limits investment model, the maximum method shows a slightly higher ENS of around 217164 MWh, this is likely because the peak values are emphasised. On 3-hourly in both investment models, the benchmark and linear interpolation methods have higher ENS values than the hourly reference which is likely because more smoothing occurs in DS. The last and maximum methods show much higher ENS around 227927 and 266256 MWh in the increased limit investment model. This may be because in these methods important fluctuations are missed and peak values are overemphasised, respectively. On 4-hourly resolution, the ENS increase is most pronounced. For the linear interpolation, maximum, median and benchmark methods the values range from 851315 to 880617 MWh in the investment model. The differences between methods are more pronounced in the increased limits investment model ranging from 239148-367636 MWh.

Overall, these results suggest that the behaviour of the down-sampled models stays similar to that in the hourly reference on 2-hourly resolution. The same holds for the down-sampled models on 3-hourly resolution, except for the linear interpolation and benchmark methods in the original investment model and the last and maximum methods in the increased-limits investment model. On the 4-hourly resolution, the only methods that stay comparable in quality to the hourly reference are the first, last, midpoint and minimum methods for the original investment model and the first, last and minimum methods for the increased limits investment model. This also indicates that on different resolutions and in the different model configurations different methods perform at better quality.

Figure 17 shows the results of the ENS in the down-sampled investment models using the hybrid method after reallocation to hourly resolution. For both the investment (top) and increased limits investment model (bottom), the ENS increases with lower resolution, which indicates that the ability of the models to capture short-term fluctuations in demand and generation diminishes and the difference with the reference becomes larger. At the 2-hourly resolution, the ENS remains close to the reference at 815102 and 203859 MWh for the original and increased limit investment model respectively, reflecting that most short-term fluctuations are still adequately captured. For the increased limits investment model, we can observe however a higher ENS at 223901 MWh for the lowest thresholds of wind onshore, suggesting that here the model struggles more at matching generation and demand likely caused by a poor reflection of short-term fluctuations. At the 3-hourly level, all methods stay close to the ENS of the hourly reference, especially for the energy demand profile experiments, while the other experiments have a slightly increase ENS, around 839224 and 212411 MWh for the original and increased limits investment models respectively. This trend persists also for the increased limits investment

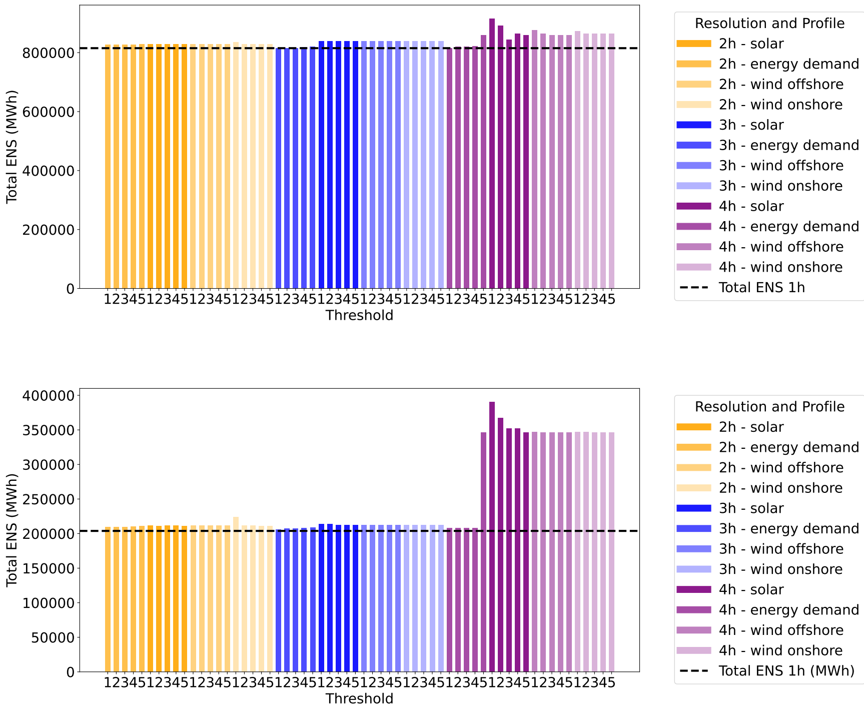


Figure 17: An overview of the energy not served after reallocation of the investment models (top) and investment model with increased limits (bottom) using the hybrid method. The energy not served is compared to that in the hourly reference model. The methods used for down-sampling are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions.

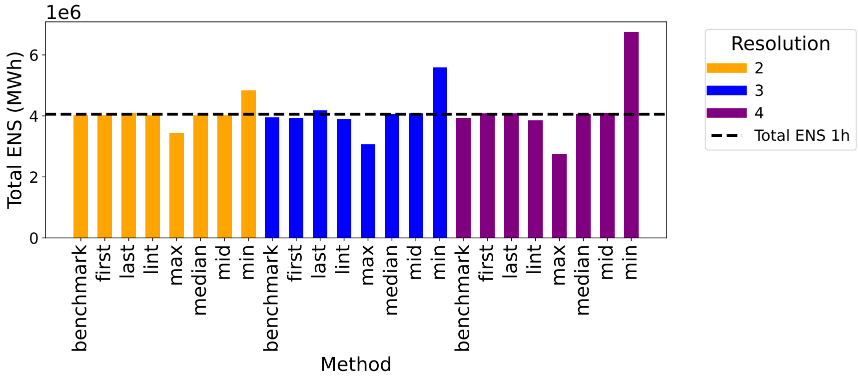
model. When moving to the 4-hourly resolution in the original investment model we see a significant increase for the highest threshold for energy demand, and the ENS in the experiments with the solar and wind profiles all range between 845064 and 915513 MWh. The increase of the ENS for the solar profile experiments is most intense. The same trend can be observed for the increased limits investment model, while here the increase in ENS is substantially higher, between 346310 and 390453 MWh, than in the original investment model.

Overall, the findings suggest that for both the original and increased limits investment models, the 2-hourly resolution generally performs well across most profiles, with minimal increases in ENS, making it a reliable choice for matching the behaviour in the hourly reference, especially for profiles like energy demand and wind. The 3-hourly resolution, while still effective, begins to show some increases in ENS, particularly in the profiles other than the energy demand. The 4-hourly resolution presents the most significant challenges in matching the hourly reference, particularly for profiles, such as solar, wind onshore and wind offshore.

4.3.2. ***Operational Models***

Figure 18 shows the results of the ENS in the down-sampled operation models using the basic statistics (top) and hybrid method (bottom) after reallocation to hourly resolution. For the basic statistics experiments, we see the ENS increases with lower resolution. Especially for the minimum method this trend is prominent, where the ENS is around 4831420 MWh, 5595561 MWh and 6748248 MWh on 2, 3 and 4-hourly resolution compared to 4056000 MWh in the reference. This indicates that the ability of this model to capture short-term fluctuations in demand and generation diminishes. For the other methods, the ENS stays close to that of the hourly reference or even drops below it in the case of the maximum around 3445930 MWh, 3068321 MWh and 2749523 MWh on 2, 3 and 4-hourly resolution, respectively. The methods that drop below the reference level, appear to have better ability to capture short-term fluctuations in demand and generation. However, the significant drop below the reference level, indicates that in these specific experiments, the behaviour of the hourly reference is not matched well, meaning that despite the lower ENS the experiments are still of lower quality.

For the hybrid method experiments, we see that on 2-hourly resolution the experiments for the energy demand, wind onshore and offshore profiles remain close to the hourly reference ENS at 4056000 MWh, and the experiments for the solar profile drop below ranging from 3779375-3993180 MWh, with an increase for higher thresholds. On 3-hourly resolution, the experiments for the energy demand, and higher thresholds for the wind onshore and offshore profiles remain close to the hourly reference ENS at 4056000 MWh but are slightly



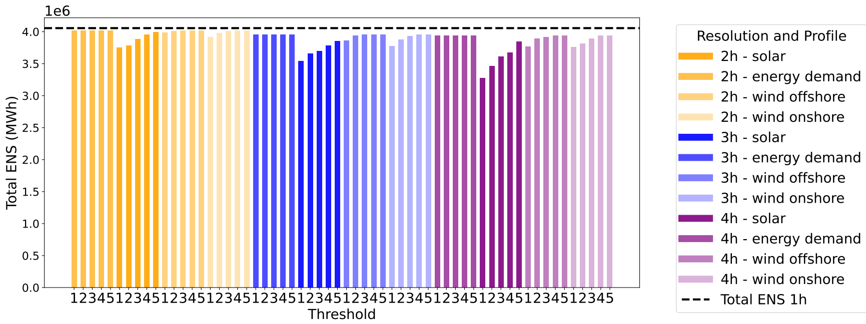


Figure 18: An overview of the energy not served after reallocation of the operational models with the basic statistics methods (top) and hybrid method (bottom). The energy not served is compared to that in the hourly reference model. The methods used for DS are displayed on the x-axis and the different resolutions are indicated with different with yellow (2h), blue (3h) and purple (4h). An increase in energy not served can be observed when moving to higher resolutions.

lower at around 3774368-3955955 MWh. Moreover, a similar decrease for higher thresholds can be observed as was the case at 2-hourly resolution for the solar profile experiments. For the solar profile, the drop is more significant ranging from 3540047-3855605 MWh, intensifying with higher thresholds. This trend persists on 4-hourly resolution but with the drop for the solar profiles even more substantial ranging from 3276391 to 3840996 MWh. Likely the reason for the decrease in the distance with the hourly reference that occurs for higher thresholds is that there appears to be a better balance between generation and demand, however, the distance between the reference level is substantial indicating that these methods perform worse in matching the behaviour of the hourly reference. Overall from these results, it becomes clear that the behaviour of the reference is best matched with the energy demand and wind profiles experiments on 2-hourly resolution and similarly on the 3 and 4-hourly resolution but for higher thresholds with the wind profiles.

4.3.3. ***Conclusion Quality Check***

The analysis of the quality check for the investment models with basic statistics methods showed that the 2-hourly resolution best preserves accuracy and aligns closely with the hourly reference in both the original and increased limits investment models with ENS around 815102 and 203859 MWh, respectively. At 3-hourly resolution, performance remains strong overall, though certain methods like linear interpolation and the benchmark in the original model, and the last and maximum methods in the increased limits model, show higher ENS around 839224 and between 227927-266256 MWh, respectively. At 4-hourly resolution, only a few methods, such as first, last, midpoint, and minimum, maintain ENS close to the benchmark. The experiments on 2-hourly resolution using the hybrid method closely match the behaviour of the hourly reference model in both the original and increased limits investment models for all profiles and thresholds. The 3-hourly resolution shows some decline, particularly for profiles other than energy demand that increase to 839224 and 212314 MWh in the investment and increased-limits investment models, while the 4-hourly resolution struggles significantly, especially with variable profiles like solar and wind where the ENS increases between 845064-915513 MWh and 347012-390452 MWh, respectively.

For the basic statistics experiments in the operational models, the minimum method showed an increased ENS of 483142, 5595561 and 6748248 MWh at 2, 3 and 4-hourly resolution. In contrast, the other methods showed an ENS close to the reference or slightly higher ranging around 4056000 MWh. This suggests that with the minimum method, the balance between generation and demand is harder to maintain. The maximum level showed a drop compared to the reference level in the range of 274952-445930 MWh across the different resolutions, suggesting that the method does not accurately match the behaviour of the hourly reference. The results for the hybrid method showed that on 2-hourly resolution the energy demand and wind profiles remained close to the hourly reference ENS, while the solar profile dropped below with values between 3779375-3993180 MWh. An increase in ENS could be observed with larger thresholds, and this trend persisted at 3 and 4-hourly resolution,

The results indicate that the energy demand and wind profiles at the 2-hourly resolution, and higher thresholds for wind profiles at the 3- and 4-hourly resolutions best match the behaviour of the hourly reference. The results of the quality check also align with the initial representation analysis and the model results presented in the previous section. The first, last and midpoint methods maintained the hourly reference behaviour in the basic statistics experiments and the energy demand profile and wind profiles experiments for the hybrid method.

4.4. Validation

This section presents the results of the validation experiments performed with climate years 1997 and 2011, for the basic statistics and hybrid DS methods. Please also refer to the results of the initial representation analysis of the 2011 and 1997 DS experiments in Appendix L.4 and Appendix L.5.

4.4.1. Basic Statistics

Figure 19, shows the results of the investment model for climate year 1997 (top) and 2011 (bottom). It can be observed that the trends in the 1997 and 2011 variations largely align with the 2008 results, in Figure 10. For the objective value or total system costs, the differences with the reference range from 0.1-15.1 % and 0.1-14.5 % for the 1997 and 2011 experiments respectively, compared to 0.1-14.9 % in the 2008 experiments. The first, last and midpoint methods outperform the benchmark with lower differences compared to the reference, similar to the 2008 experiments. For the gas investment decisions, there is a similar trend in the 1997 results compared to 2008. The first last and midpoint methods on 2 and 4-hourly resolution, show differences ranging from 0.5-0.7 % and 0.1-2.7 %, respectively compared to 0-0.2 % and 0.8-1.7 % in the 2008 experiments. When

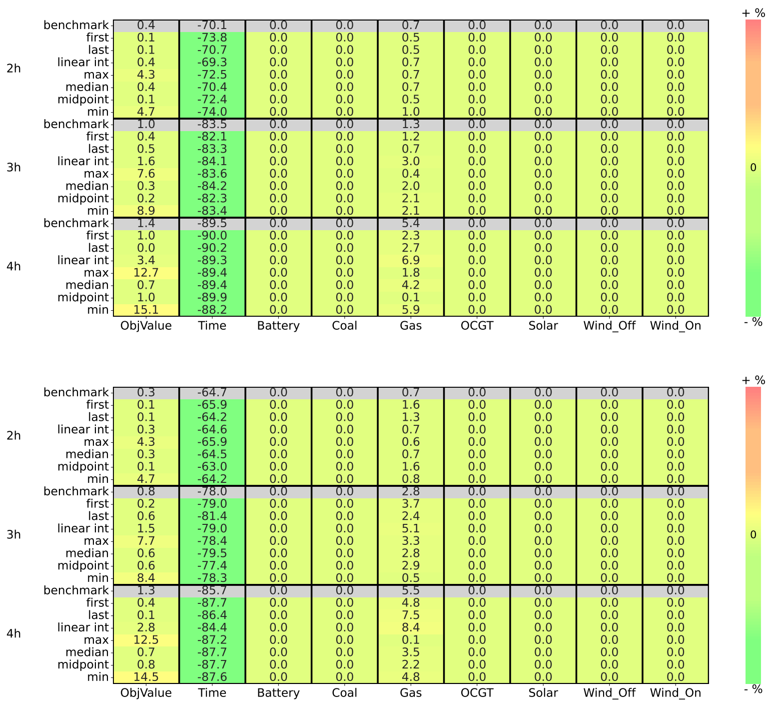


Figure 19: Overview of the results of the validation of the basic statistics methods in the investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not.

Figure 20: Overview of the results of the validation of the basic statistics methods in the increased limit investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not.

looking at the minimum and maximum methods, the performance is variable similar to the 2008 results. On 2 and 3-hourly resolutions the differences with the reference remain which is relatively low ranging from 0.4-2.1 % compared to 0.5-1 % in the 2008 experiments. However, while in the 2008 experiments, the maximum method only outperformed the benchmark on 3-hourly resolution, the maximum method outperformed the benchmark in the 1997 results on all resolutions for the gas investment decisions. In the 2011 results, a slightly different trend can be observed as well, where the differences with the reference are higher for methods like the first last and midpoint ranging from 1.3-7.5 % compared to 0-1.3 % in the 2008 experiments. Generally, the experiments struggle to outperform the benchmark, this is especially the case for methods that previously stably outperformed the benchmark. The time reduction in all experiments increases upon lower resolution. However, the reduction in the 1997 experiments ranging from 70.1-90.2 % appears to be more pronounced than in the 2008 and 2011 experiments, with reductions ranging from 66.2-90.6 % and 63.0-87.7 %. The solve times of the hourly references are 146.4 and 132.4 seconds in the 1997 and 2011 models, respectively. Overall, the performance of the promising methods in the 2008 experiments: the first, last and midpoint methods seem to be sensitive to input data, especially on 2 and 3-hourly resolution for the gas investment decisions.

Figure 20, presents the results of the investment model with increased limits for the climate year 1997 (top) and 2011 (bottom). For the objective value or total system costs, the trends in the 1997 and 2011 variations align with the 2008 results with differences from the reference ranging from 0.1-28.1 % and 0.2-25.1%, respectively, compared to 0.0-27.5 % in 2008, see Figure 11. Moreover, the methods that consistently outperform the benchmark with lower differences compared to the reference are again the first, last and midpoint methods. However, for the investment decisions, there is a difference in trend from the 2008 results. In the 1997 results, instead of investments in coal like in the 2008 results, investment decisions are made in OCGT. For the 2011 variation, investments in coal can be observed, however, the difference with the hourly reference ranging from 0-18.3 % appears to be larger than in the 2008 model ranging from 0-9.9 %. The midpoint, last and first methods perform more variably across different resolutions in both 1997 and 2011. For the wind offshore investment decisions, the performance of these methods largely aligns with the 2008 results. The trend in gas investments also still largely aligns with the 2008 results. However, the first method performs more poorly on the 2 and 3-hourly resolution in the 1997 results and on the 3-hourly resolution in the 2011 results, with differences compared to the hourly reference that are higher than the difference of the benchmark. For the midpoint method, the performance is also slightly different, in the 2008 model it outperformed the benchmark consistently across all resolutions for the gas investment. While this is still the case in the 2011 results, in the 1997 results on the 2 and 4-hourly resolution it fails to outperform the benchmark. For the investment decisions in coal, we can see a similar trend in both the 2011 and 2008 results, where the linear interpolation method performs poorly, while generally the other methods perform close to the benchmark difference with the reference or beat it. Overall, the first, last and midpoint methods seem to be sensitive to input data, however, they still, though not consistently across all resolutions, beat the benchmark. Additionally, across both the 1997 and 2011 experiments a similar reduction in solve time can be observed upon lowering the resolution. The solve time reduction ranges from 67.3-91.8 % in 1997 and 72.0-92.7 % in 2011 compared to 67.8-91.0 % in 2008. The solve times of the hourly references are 120.3 and 122.2 seconds in the 1997 and 2011 models, respectively. Thus despite their sensitivity to the input data, the first last and midpoint methods still come out as promising methods when the goal is reducing solve time while achieving a lower difference with the reference model than the benchmark on all metrics.

Figure 21, presents the results of the operational models for climate year 1997 (left) and 2011 (right). For both the 1997 and 2011 results, the trend in objective value or total system cost matches the 2008 results in Figure 12. The differences range from 0.6-13.3 % and 0.7-13.4 % in the 1997 and 2011 experiments, respectively compared to 0.3-14.1 % in the 2008 experiments. The maximum and minimum methods exhibit differences larger than the benchmark and the first last and midpoint methods outperform the benchmark with lower differences. For the flows, the trend in the performance is similar to the 2008 results, the differences are all between 18 and 29 % and remain stable across different resolutions. No single method consistently outperforms the benchmark and their performance is generally close to that of the benchmark. In terms of net position, the trends are significantly different, for the 1997 and 2011 experiments the differences range from 21.2-311.7 % and 9.4-97.9 %, respectively compared to 11.2-116.2 % in the 2008 experiments. Additionally, the differences on 3-hourly resolution in the 1997 experiments are significantly higher than on 2 and 3-hourly resolution, which is not the case in the 2008 and 2011 experiments. Though overall in the 2008 results and both the 1997 and 2011, it appears to be easier to outperform the benchmark difference with the reference. For the 2008 model, this is especially the case on the 4-hourly resolution while for the 1997 and 2011 models, this is the case at 3 and 2-hourly resolution, respectively. In all cases, the methods that are performing well, are consistently the first, last and midpoint methods, although the performance varies across different resolutions. Again a similar trend in solve time reduction can be observed across the different models, in the 1997 and 2011 experiments, the drop in solve time ranges from 65.5-89.8 % and 48.9-90.6 %, compared to 56.0-86.4 % in 2008. The solve times of the hourly references are 23.0 and 25.9 seconds in the 1997 and 2011 models, respectively. Thus, when using different input data it remains challenging to outperform the benchmark accuracy in terms of generation flows, but the ability to manage net position improves for methods such as the median and midpoint.

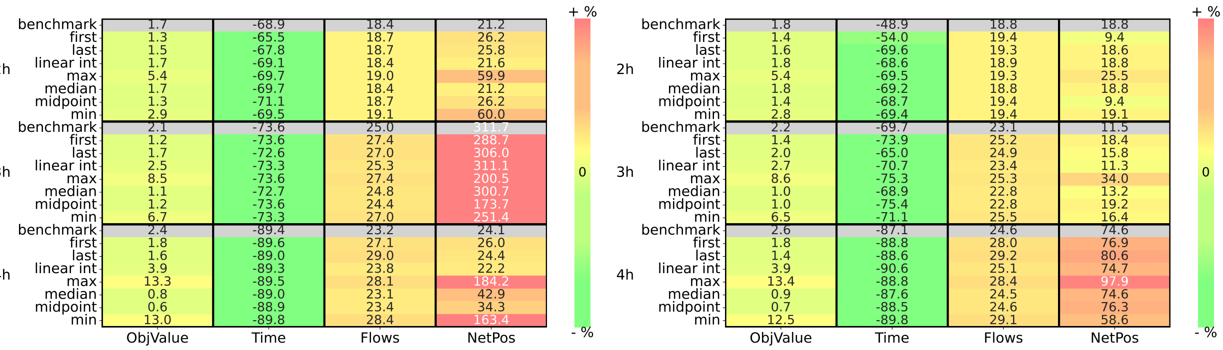


Figure 21: Overview of the results of the validation of the basic statistics methods in the operational model with climate year 1997 (top) and 2011 (bottom). On the x-axis the evaluation metrics are displayed including the objective value (total system cost), solve time, flows and net position. On the y-axis, the different down-sampling methods are displayed including the different resolutions. The numbers in the cells of the matrix indicate the percentage difference between the experiment and the hourly reference model. The benchmark is indicated in grey. Note that for the objective value flows, and net position the differences are absolute while for the solve time, they are not.

4.4.2. ***Hybrid***

Figure 22, presents the results of the investment model with climate years 1997 (top) and 2011 (bottom). Again for the objective value or total system cost, the trends with the 2008 model results in Figure 13 align. The differences range from 0.4-24 % and 0.3-22 % in the 1997 and 2011 experiments compared to 0.5-22 % in the 2008 experiments. Overall, the experiments struggle to outperform the benchmark and mostly show higher differences compared to the reference. Higher thresholds for the wind offshore and energy demand profiles show differences close to that of the benchmark across all resolutions. Moreover, like in the 2008 results, higher resolutions tend to exhibit lower differences with the reference.

The trend in how large the differences are compared to the reference also aligns across the different climate years. The energy demand and solar profiles have substantially higher values than the wind profiles across all resolutions in both the objective value and investment decisions. Specifically, for the investment decisions in the 2008 results, on 2-hourly resolution higher thresholds for energy demand, solar and wind onshore profiles outperformed the benchmark with lower differences around 0.1-0.6 % compared to 0.7 % for the benchmark. For the 1997 results, the trends largely align. The highest thresholds for the energy demand profile show lower differences than the benchmark, and the highest thresholds for solar and wind profiles perform at the benchmark difference of around 0.7 %. However, in the 2011 results, no experiments outperformed the benchmark, but the scores are very close to the benchmark around 0.7 %. On 3-hourly resolution, the trend is similar across all three years, where most methods have higher differences than the benchmark and higher thresholds for wind profiles perform close to the benchmark around 1.3 % for the 1997 experiments. For the 2011 results the intermediate and higher thresholds for wind offshore and onshore, respectively, slightly outperform the benchmark with differences below 2.8 %. On the 4-hourly resolution, the trends are most deviant. The higher thresholds for the energy demand perform with a lower difference than the benchmark between 0.4-5.4 % in the 1997 results, but not in the 2011 and 2008 model results. Moreover, the highest threshold for the solar method performs well in the 2008 model while this is not the case for the 2011 and 1997 models. A similar trend in solving time reduction can be observed across the different models. In the 1997 and 2011 experiments, the drop in solve time ranges from 65-90 % and 58-87 %, compared to 66-91 % in 2008. The solve times of the hourly references are 120.3 and 122.2 seconds in the 1997 and 2011 models, respectively. Overall, the trends for the investment model in 2008 largely align with those of the 2011 and 1997 model results. When attempting to match the investment decisions on 4-hourly resolutions the methods appear most sensitive to the input data. The methods that still come out promising are the energy demand profile and wind profiles.

Figure 23, presents the results of the investment model with increased limits for climate years 1997 (top) and 2011 (bottom). Again for the objective value the trends with the 2008 model results in Figure 14 align. The differences range from 0.5-20 % and 0.1-24 % in the 1997 and 2011 experiments compared to 0.3-23 % in the 2008 experiments. In all climate years, only the higher thresholds for the energy demand exhibit lower differences than the benchmark. Higher resolutions tend to exhibit lower differences with the reference. For

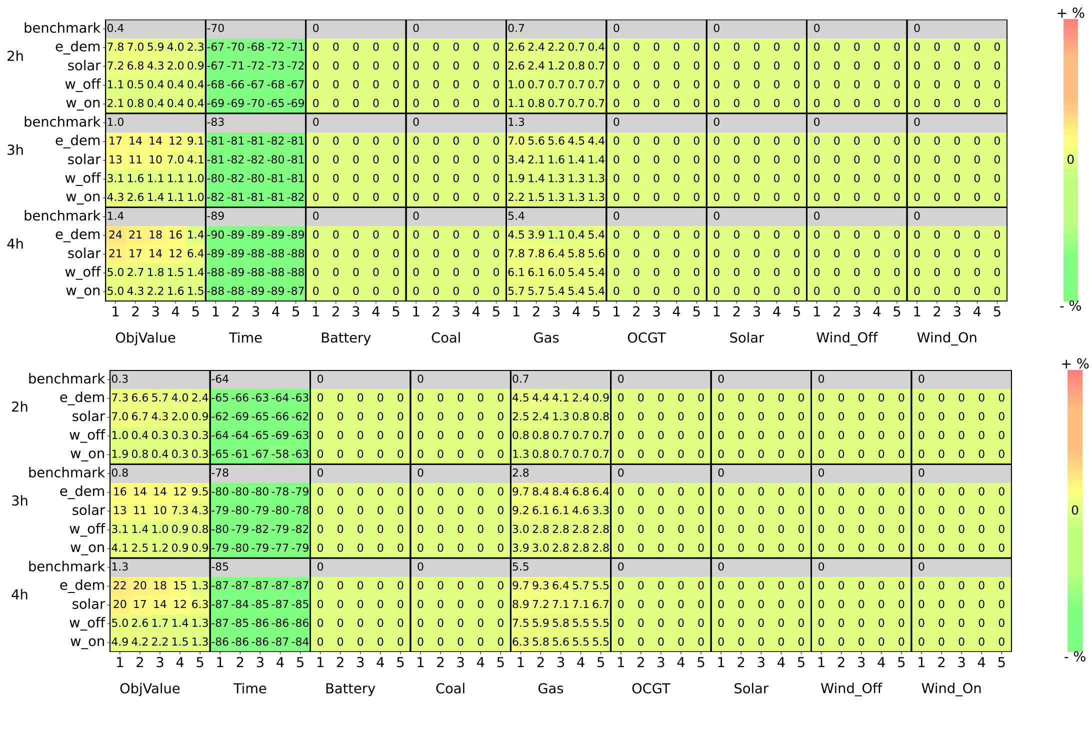


Figure 22: Overview of the results of the validation of the hybrid methods in the investment model with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not.

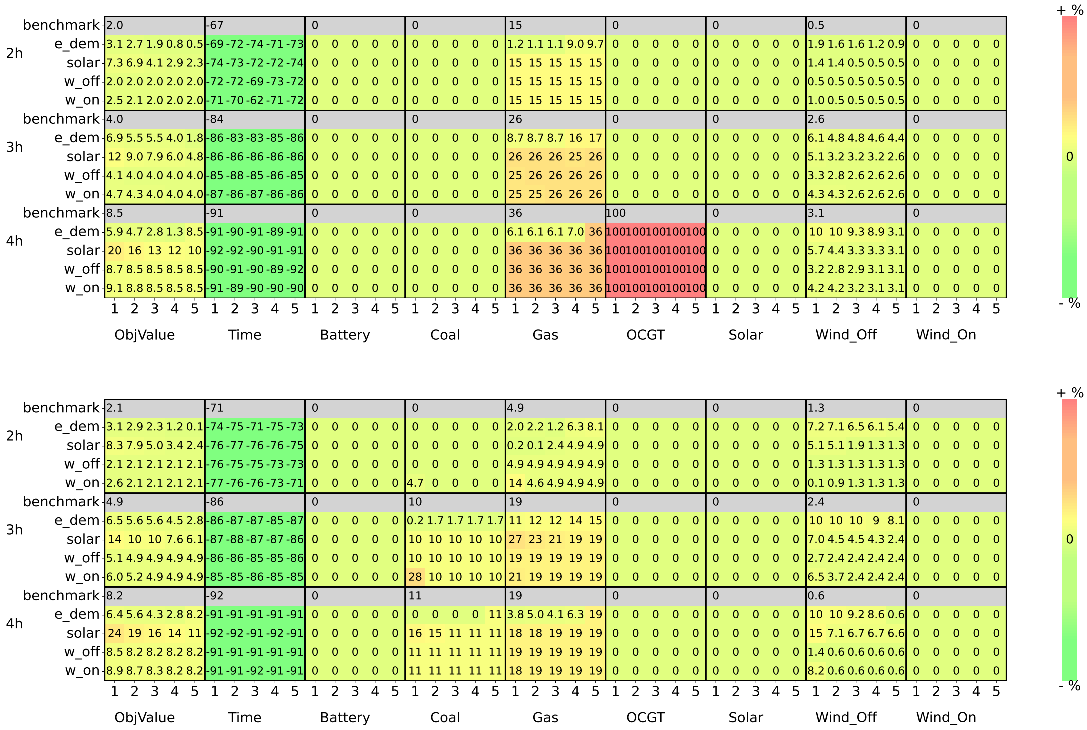


Figure 23: Overview of the results of the validation of the hybrid methods in the investment model with increased limits with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the investment decisions for generation and flexibility assets in the model. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value and investment decisions the differences are absolute while for the solve time, they are not.

the investment decisions in wind offshore, the trends for solar, and the wind profiles are largely in line for the different years, where the solar profile and the lower thresholds for the wind profiles consistently show differences with the reference larger than the benchmark, and the higher thresholds perform similar to the benchmark. In the 2008 results, the energy demand profiles outperformed the benchmark on 2 and 3-hourly resolution however this is not the case for the 1997 and 2011 results. Regarding the investment decisions in gas, the methods appear to perform more stable, though with larger differences with the reference ranging from 1.1-36 % in 1997 and 0.1-19 % in 2011 compared to 20-59 % in 2008. In all models, most thresholds for the energy demand profile perform well with lower differences than the benchmark consistently over all resolutions. The same can be observed for the solar profile and lower thresholds for wind onshore on the 4-hourly resolution. On the 3-hourly resolution, however, more methods appear to show lower differences with the reference than the benchmark in the 2008 model. For the coal investments in the 2008 model, the effect was only observed on the 4-hourly resolution but for the 2011 model, the effect can also be observed on 2 and 3-hourly resolution. Again a similar trend in solve time reduction can be observed across the different models, in the 1997 and 2011 experiments, it ranges from 62-92 % and 71-92 %, compared to 66-92 % in 2008. The solve times of the hourly references are 120.3 and 122.2 seconds in the 1997 and 2011 models, respectively. Overall, the experiments with energy demand and wind profiles still come out strong, outperforming the benchmark and steadily performing at the benchmark accuracy, respectively.

Figure 24 presents the results of the operation model for climate years 1997 (top) and 2011 (bottom). Again for the objective value the trends align with the 2008 model results in Figure 15. The trend in the performance of the experiments in flows is similar to the 2008 results for both 1997 and 2011 results. The percentage differences compared to the hourly reference remain quite stable across all resolutions and the difference in performance between different profiles is minimal all ranging between 18-24 %. The net position trends across the three years are significantly different, where in the 2008 experiments the differences observed ranged from 13-115 %, and in the 1997 and 2011 results they ranged from 14-503 % and 10-90 %. Notably, where the differences on 3-hourly resolution were lower than on 2- and 4-hourly resolution in the 2008 and 2011 results, in the 1997 results they were significantly higher ranging from 178-503 %. On the 2-hourly resolution in the 2008 results the higher thresholds for the wind profiles outperformed the benchmark, while for the 1997 results, this happens for the higher thresholds of the energy demand and solar profiles and in the 2011 results for

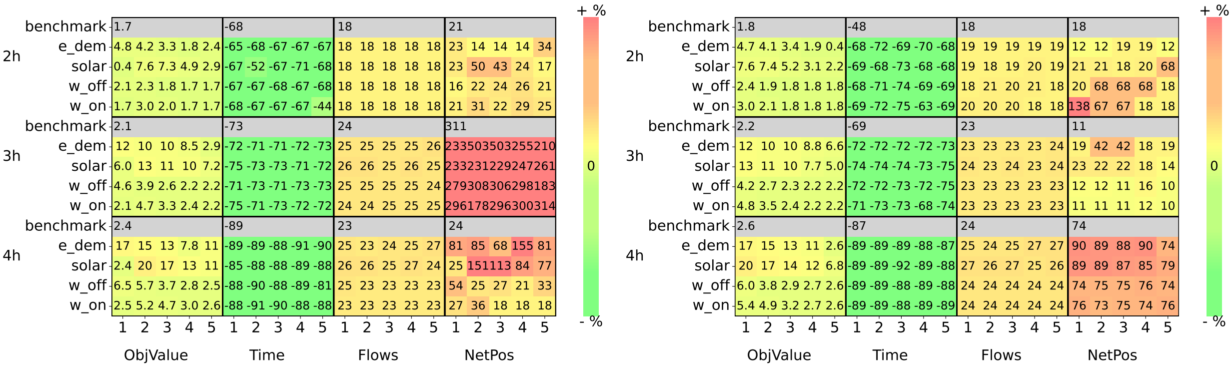


Figure 24: Overview of the results of the validation of the hybrid methods in the operational model with climate year 1997 (top) and 2011 (bottom). On the x-axis, the different thresholds used for the hybrid method are indicated with a number from 1 to 5. The threshold is ranked from low (1) to high (5), and Table G.10 summarises the exact threshold per profile. The x-axis is further subdivided into the evaluation metrics including the objective value, solve time and the flow and net position behaviour. On the y-axis, the different profiles that were down-sampled using the hybrid method are displayed including the different resolutions. The benchmark is indicated in grey. Note that for the objective value, flows and net position the differences are absolute while for the solve time, they are not.

energy demand and wind onshore profiles. On the 3-hourly resolution for the 1997 model, almost all methods outperform the benchmark and for the 2011 model, only higher thresholds for wind offshore and onshore do so, while in the 2008 model, none of the methods were able to beat the benchmark. On the 4-hourly resolutions, the trends are more or less similar, though in the 2008 model energy demand profile performed at the benchmark accuracy and thresholds for the wind profiles outperformed it. In the 1997 results, we see also the wind profiles outperforming the benchmark, similarly for the wind onshore profiles and the highest thresholds for the energy demand profile in the 2011 models. Again similar trend in solve time reduction can be observed, in the 1997 and 2011 experiments, the drop in solve time ranges from 44-91 % and 48-92 %, compared to 66-89 %. The solve times of the hourly references are 23.0 and 25.9 seconds in the 1997 and 2011 models, respectively. Overall, the results appear highly sensitive to the input data, especially for the net position metrics, where the trend is most deviant in the 2011 and 1997 models from the 2008 model results.

4.4.3. Conclusion Validation

Basic Statistics

The validation experiments for the investment model show that the trends observed in the 2008 results are similar to the 2011 and 1997 results. When approximating the accuracy of the objective value the first, last and midpoint methods consistently outperform the benchmark However, for matching the accuracy in gas investment decisions on 2- and 3-hourly resolutions these methods show sensitivity to input data. This is especially the case in the 2011 results where it is difficult to outperform the benchmark. The performance of the minimum and maximum methods is similar to the 2008 results. They generally show high differences with the reference, although in 1997 the maximum method performs better across all resolutions. In the increased limits investment model, the 2011 and 1997 trends align with the 2008 model when matching the objective value. However, the impact on investment decisions varies, with a shift from investment coal to OCGT in the 1997 model. When approximating the accuracy of the hourly resolution of the midpoint, last and first methods show sensitivity to the input data, resulting in different performances across the resolutions. Nonetheless, these methods generally outperform the benchmark. For the operational model results the trends in 1997 and 2011 also generally align with the 2008 results. This is particularly the case when approximating the objective value and generation flows on hourly resolution. Most methods perform similarly to the benchmark. For the net position, it is easier to outperform the benchmark, with the 1997 model performing best at the 3-hourly resolution and the 2011 model at the 2-hourly resolution. The first, last and midpoint methods consistently perform well, and unlike in 2008, the median and midpoint methods in 1997 and the median method in 2011 consistently outperform the benchmark across all metrics at different resolutions.

Hybrid

In the investment model the trends in the 2008 results are similar to those in the 1997 and 2011 results when approximating the objective value. This is also the case for the overall performance of higher thresholds for wind offshore and energy demand profiles. These methods, however, appear to be sensitive to input data when approximating the investment decisions in the reference on 4-hourly resolution. Nonetheless, the methods that still come out strong are the energy demand profile and wind profiles. In the increased limits investment model, the methods seem most sensitive to the input data when attempting to match the investment decisions for wind offshore. Despite this, the experiments with energy demand and wind profiles still come out strong, outperforming the benchmark and steadily performing at the benchmark accuracy, respectively. Lastly, for the operational model, the experiments show varying performance across the climate years 1997, 2011, and 2008. The trend for the objective value is similar to the 2008 model, however, the performance in flows differs, with more methods outperforming the benchmark in 1997 and fewer in 2011. The results indicate that the hybrid experiments are sensitive to input data, particularly for the net position, where the trends in 1997 and 2011 differ most from the 2008 results.

5. Discussion

The primary objective of this research was to assess how effective different DS methods are in reducing the temporal resolution in energy system optimisation models while approximating accuracy at hourly resolution and reducing complexity. This objective was addressed in three phases, of which the first comprised an initial representation analysis to assess how well the different DS methods represent the original data in terms of energy value distribution and ramp distribution. In the second phase, the down-sampled input profiles were applied within an operational, an investment and an increased limits investment model configuration in a Northwestern Europe case study. After the second phase the impact of the different DS methods on the model accuracy was assessed by evaluating based on the objective value or total system costs, investment decisions, flows and net position behaviour. The impact on complexity reduction was assessed by evaluating the model solve time. Moreover, the impact of the different methods on model accuracy was compared to the benchmark method that uses averaging for aggregation.

The main findings demonstrate that regardless of the DS method, aggregating the input profiles resulted in a significant reduction in computational complexity. With a reduction across the different model configurations ranging from 44-75 % on 2-hourly resolution, 62-88 % on 3-hourly resolution and 81-92 % on 4-hourly resolution. In terms of model accuracy, no DS method consistently outperforms the benchmark method on all resolutions and in all model configurations. Nonetheless, a trend is recognisable, especially for the basic statistics methods where the first, last and midpoint methods often outperform the benchmark, showing lower differences with the hourly references. In contrast, the maximum and minimum methods showed consistently lower performance than the benchmark, with higher differences with the hourly reference. For the hybrid method, a trend could be discerned as well, where the DS of the energy demand profiles often resulted in lower differences with the reference than the benchmark method and DS wind profiles resulted in performance close to the benchmark. The experiments with solar profiles generally exhibited higher percentage differences with the reference compared to the benchmark.

The outcomes show however that the impact of the different methods is highly dependent on the model configuration, which aligns with findings from [15] that investigate the effect of time series aggregation methods on optimal energy system design. To illustrate, the first last and midpoint methods performed stably across all configurations when considering the objective value in all model configurations, with differences compared to the hourly reference of 0.1-1.7 % generally lower than the benchmark. Moreover, a similar trend was observed in the investment decisions in both variations of the investment configurations, where the methods exhibited differences with the hourly reference ranging from 0.0-1.7 % for gas investment decisions in the investment configuration and 4.1-39.7 % and 0.2-1.8 % for gas and wind offshore investment decisions in the increased limits configuration. Evaluating the flow behaviour in the operational configuration showed that, while the differences with the hourly reference remained stable ranging from 19-28% it appears more difficult to outperform the benchmark, a similar trend occurred when evaluating the flows in the investment models after reallocation to hourly resolution with differences ranging around 11 and 27 % for the investment and increased limits investment configurations, respectively. Moreover, the low difference in investment decisions and flows in the operational and investment configurations resulted in high differences in net position compared to the reference, generally ranging from 11-823 %. While it is challenging to outperform the benchmark accuracy in terms of generation flows, especially at lower resolutions, the ability to manage net position improves with methods that focus on specific points or extremes, such as the first, last and midpoint methods. For the hybrid method mostly the higher thresholds for the energy demand profiles showed the ability to outperform the benchmark in terms of objective value exhibiting differences with the reference ranging from 0.1-22%, and experiments with wind profiles generally stayed close to the benchmark between 0.5-11 %, in both the operational and investment configurations. A similar trend could be observed for energy demand and wind profiles experiments in the gas investment decisions in the investment configuration and the wind offshore investment decisions in the investment configuration with increased limits. Although their ability to outperform the benchmark decreased upon lowering the resolution. For the gas investment decisions in the increased limits configuration, the trends observed are different as well, where generally the lower thresholds for the energy demand performed with lower differences with the reference compared to the benchmark and the differences are generally higher than in offshore wind investment decisions. Both of these observations are likely because of the role of the gas assets in addressing peak demand. When considering the flow behaviour similar trends occur as did for the basic statistics methods where the difference between the methods is low and they all exhibit a difference with the reference around 11% in the investment configurations after reallocation to hourly resolution and around 21 % in the operational configuration. It also appears more difficult to outperform the benchmark for this metric. The low difference in investment decisions and flows in the operational and investment configurations resulted in high differences in net position compared to the reference generally ranging from 409-823 %.

The trends observed in the performance of the different methods can be linked to their performance in the representation of energy values and ramp distributions. The results of the initial representation analysis showed that the first, last and midpoint methods exhibited low CAQE and CAQE ramp errors and outperformed the benchmark. In contrast to, for example, the minimum and maximum methods that consistently showed substantially higher CAQE but in some cases improved CAQE ramp errors, similar to what [20, 21, 25] show when using minimum and maximum methods for better extreme value and fluctuation representation. In the hybrid method, it is more difficult to exactly link the performance to the results of the initial representation analysis. Generally, intermediate and higher thresholds showed better balances between ramp and energy value representation also similar to the findings in [20, 21, 25], however in most cases the representation in energy values was better in the benchmark method. Moreover, in some cases lower thresholds were able to outperform the benchmark, suggesting that a good balance between energy value and ramps is not the only requirement for improved model performance. The difference in behaviour between the operational and investment configurations in both the basic statistics and hybrid methods can be attributed to the focus of each configuration. The difference in results and behaviour of the investment and operational model configurations is because of their different objectives. The goal in the investment model is long-term planning and in this context, it is more important to ensure that there is enough capacity to meet demand in the future. For the model to accurately do this is crucial that the extremes in the input profiles are captured [13]. This likely explains why more of the DS experiments can outperform the benchmark in the investment configurations. On the other hand, the objective of the operational model is to manage a real-time balance between supply and demand and in this context preserving short-term fluctuations is important [13].

These results imply that the first, last and midpoint methods generally provide a better alternative than the benchmark when reducing the temporal resolution while reducing model complexity and approximating the hourly resolution accuracy in the objective value. This is similar for the higher thresholds used in the energy demand profile and wind profile experiments in the hybrid experiments although less consistent. On the other hand, in the operational configuration, it is more difficult to retain accuracy in terms of flow behaviour and net position behaviour however, the alternative methods generally perform better than the benchmark method with lower differences compared to the hourly reference.

Several limitations should be acknowledged when interpreting these findings. Firstly, the results of the hybrid method showed to be sensitive to the thresholds that were used in the DS process. The threshold choices were done based on the range identified for each specific profile, and by dividing the range in uniform steps to explore the threshold range. Moreover, a grid search optimising for the ramp representation in the down-sampled profile was used to determine an optimal threshold, whereas the model performance showed to be dependent not only on the ramp representation. Overall, experiments with higher thresholds showed better performance, likely because of the better balance between ramp and energy content representation. Nonetheless, the threshold choice can be improved and thereby the performance, by further exploring the threshold range and optimising for additional characteristics such as energy value distribution and extreme value representation in the grid search. Moreover, the determination of the threshold range was based on the Dutch energy demand, solar and wind profiles and those thresholds were imposed on the profiles of the other countries in the Northwestern Europe case study. This may have affected the energy value distribution and ramp distribution when DS the profiles from the other countries. To add to this, the initial representation analysis in Section 4.1.1 was also based on the Dutch profiles. Nevertheless, the overall trend in performance of the basic statistics and hybrid experiments in model metrics remained consistent and the trends could still be linked with the initial representation analysis.

In addition, the validation with different climate years indicated that the results of the DS experiments are sensitive to input data. Generally, the basic statistics methods showed to be more robust with fewer differences from the original 2008 climate year experiments. The first, last and midpoint methods remained the most promising, often beating the benchmark, though not always consistently across all resolutions and all model configurations. The hybrid methods were shown to be sensitive to the input data. Overall, the performance of the experiments was variable however the general trends remained, such as the generally better scores for higher thresholds, and the energy demand profile and wind profile experiments beating the benchmark regularly and steadily performing at the benchmark accuracy, respectively. The variation is likely also because the thresholds used were those tailored to the 2008 climate year data, which may have affected the performance in the 2011 and 1997 experiments. Thus considering this, the outlook for the hybrid methods in reducing temporal resolution while approximating the accuracy at hourly level in energy system optimisation models remains promising.

Logically, when lowering the resolution of the input profiles regardless of the methods used for DS the solve time would remain similar. However, a level of variability of the solve time could be observed within the experiments for the different resolutions. For example, the solve time of the hybrid experiments in the operational model configuration ranged from -68 to -44 %. This was likely because of resource contention, affecting the solve time. In future work this should be addressed by performing each experiment e.g. 10 times, to investigate the solve time distribution. Lastly, it must be considered that the generalisability of the results remains limited to the North Western European case study and although the solve times of hourly references are short ranging from 23-146.4 seconds the observed decrease in complexity highlights the potential for significant reductions in solve time in models on larger scales. Despite the limitations, this study underlines the potential of alternative DS methods to reduce the temporal resolution while approximating the accuracy on hourly level in the objective value and investment decisions compared to the benchmark method. Future work could expand on these experiments by applying them on a larger scale to evaluate the effect on complexity reduction. Furthermore, for the hybrid methods combinations of down-sampled profiles in models can be explored, allowing to harbour the strengths of each profile and have a combinatorial effect on improving the model accuracy approximation. Moreover, combinations of different basic statistics methods could be considered in the hybrid method, such as using the first last minimum and midpoint in combination with the benchmark averaging, similarly to [23] that extend Adaptive Piece-wise Aggregate Approximation to include both minima and maxima next to averaging to represent financial time series. The results of this study are also meaningful for applications that currently mainly use averaging to down-sample for example when DS of typical periods [27, 28]. Moreover, in the mathematical formulation of Tulipa, no ramping constraints are present, and it must be emphasised that this may have affected the model performance of DS methods that present ramps better compared to the benchmark. Therefore, future work could also explore the application of alternative DS methods in energy system optimisation models that include ramping constraints. Finally, a deeper investigation into the optimal thresholds used for specific profiles in the hybrid methods would be a promising extension of this research.

6. Conclusion

In conclusion, this research aimed to assess which DS methods are most effective in reducing the temporal resolution of time series to reduce model complexity while approximating the model accuracy at hourly resolution. This was done with a Northwestern European case study using an operational model configuration and two variations of an investment model configuration. The ability of methods to approximate model accuracy on hourly resolution was evaluated by comparing the objective value, investment decisions, flows and net position in the experiments to the hourly reference. The main findings show that regardless of the method, DS resulted in a significant reduction in solve time and thus computational complexity. While no method consistently showed lower differences with the hourly reference than the benchmark, a clear trend is recognisable. This is especially the case for the basic statistics methods where the first, last and midpoint methods often did show lower differences, particularly in approximating accuracy in total system cost or objective value and investment decisions. In contrast, the maximum and minimum methods consistently showed higher differences with the hourly reference for these metrics. The hybrid method also shows potential, especially when DS the energy demand profiles resulting in lower differences in objective value and investment decisions compared to the reference than the benchmark method. DS wind profiles resulted in performance close to the benchmark. The experiments with solar profiles generally exhibited higher percentage differences in objective value and investment decisions with the reference compared to the benchmark often.

The trends observed in the ability of different DS methods to approximate the model accuracy in objective value and investment decisions on hourly resolution are linked to their ability to represent energy values and ramp distributions. The first, last and midpoint methods exhibited lower CAQE and CAQE ramp errors compared to the original data than the benchmark. In contrast, the minimum and maximum methods consistently showed significantly higher CAQE than the benchmark, though in combination with lower CAQE ramp errors. For the hybrid method, the link with the initial representation analysis is less direct, generally intermediate and higher thresholds showed a better balance, with low errors in ramp and energy value representation. This suggests that a good balance between energy value and ramps is not the only requirement for an improved approximation of model accuracy on hourly resolution. In the operational configuration, similar trends could be observed in the approximation of accuracy for the objective value, however for the flow and net position the differences with the hourly reference remained much higher. This difference in behaviour can be attributed to the different objectives of these model configurations. The goal of the investment model is long-term planning to ensure enough capacity to meet demand in the future. Here the extremes in the input profiles must be captured [13]. The objective of the operational model is to manage a balance in real-time between supply and demand and preserving short-term fluctuations is important [13]. This likely explains why more of the DS experiments can outperform the benchmark in accuracy approximation in the investment configurations.

The results imply that the first, last and midpoint methods generally provide a better alternative than the benchmark when reducing the temporal resolution to reduce model complexity while approximating model accuracy for the objective value and investment decisions. This is similar for the higher thresholds for energy demand profiles and wind profiles in the hybrid experiments although less consistently. These findings contribute to the understanding of the relationship between energy value distribution and ramp representation and model performance. They provide insight into the impact of alternative DS methods on reducing model complexity while approximating accuracy and emphasise their benefit in this subtext compared to the benchmark method. To better understand the implications of these results, future research should address further exploring the threshold ranges for the hybrid method and optimising for additional characteristics such as energy value distribution and extreme value representation in the grid search. Moreover, the application of these methods on a larger scale can be explored to extend the results. Furthermore, for the hybrid methods using combinations of down-sampled profiles in models can be explored, allowing to harbour the strengths of each profile and have a combinatorial effect on improving the model accuracy and combinations of different basic statistics methods could be considered in the hybrid method, such as using the first last and midpoint in combination with the benchmark averaging. Lastly, the results of this study are also meaningful in applications that currently mainly use averaging to down-sample for example when DS of typical periods [27, 28] or in the application of optimisation models that include ramping constraints.

Overall, the findings highlight the potential of the alternative DS methods to provide a more successful option than the benchmark using averaging in reducing complexity while approximating accuracy on hourly resolution. The methods for applying alternative DS methods may potentially be applied, in the future, in other optimisation contexts beyond energy system optimisation where time series data is crucial and computational complexity needs to be reduced without sacrificing accuracy.

7. Acknowledgements

I would like to express my sincere gratitude and appreciation to my supervisors, dr. Matteo Gazzani and Zhi Gao, MSc. Not only have they provided support, insightful feedback, and valuable suggestions during our weekly meetings, but they have also made our work together incredibly enjoyable. I would also like to thank dr. Germán Morales-España for his guidance feedback and expert opinion in the process and dr. Diego Tejada Arango for helping with setting up the Tulipa environment and handling the server environment for running the experiments. Their enthusiasm and positive attitude have truly enhanced the collaborative experience at TNO.

References

[1] G. Morales-España, L. Ramírez-Elizondo, B. F. Hobbs, Hidden power system inflexibilities imposed by traditional unit commitment formulations, Applied Energy 191 (2017) 223–238. doi:10.1016/j.apenergy.2017.01.089.
URL https://linkinghub.elsevier.com/retrieve/pii/S0306261917301009

[2] N. Sarajpoor, L. Rakai, J. Arteaga, N. Amjady, H. Zareipour, Time Aggregation in Presence of Multiple Variable Energy Resources, IEEE Transactions on Power Systems 39 (1) (2024) 587–601. doi:10.1109/TPWRS.2023.3236842.
URL https://ieeexplore.ieee.org/document/10016747/

[3] C. D. Iweh, S. Gyamfi, E. Tanyi, E. Effah-Donyina, Distributed Generation and Renewable Energy Integration into the Grid: Prerequisites, Push Factors, Practical Options, Issues and Merits, Energies 14 (17) (2021) 5375, number: 17 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/en14175375.
URL https://www.mdpi.com/1996-1073/14/17/5375

[4] S. Buchholz, M. Gamst, D. Pisinger, A comparative study of time aggregation techniques in relation to power capacity expansion modeling, TOP 27 (3) (2019) 353–405. doi:10.1007/s11750-019-00519-z.
URL http://link.springer.com/10.1007/s11750-019-00519-z

[5] C. Brunner, G. Deac, S. Braun, C. Zöphel, The future need for flexibility and the impact of fluctuating renewable power generation, Renewable Energy 149 (2020) 1314–1324. doi:10.1016/j.renene.2019.10.128.
URL https://www.sciencedirect.com/science/article/pii/S096014811931626X

[6] Morales-Espana, J. M. Latorre, A. Ramos, Tight and Compact MILP Formulation for the Thermal Unit Commitment Problem, IEEE Transactions on Power Systems 28 (4) (2013) 4897–4908. doi:10.1109/TPWRS.2013.2251373.
URL http://ieeexplore.ieee.org/document/6485014/

[7] M. Daadaa, S. Séguin, K. Demeester, M. F. Anjos, An optimization model to maximize energy generation in short-term hydropower unit commitment using efficiency points, International Journal of Electrical Power & Energy Systems 125 (2021) 106419. doi:10.1016/j.ijepes.2020.106419.
URL https://www.sciencedirect.com/science/article/pii/S0142061519342218

[8] C. K. Chyong, D. Newbery, A unit commitment and economic dispatch model of the GB electricity market – Formulation and application to hydro pumped storage, Energy Policy 170 (2022) 113213. doi:10.1016/j.enpol.2022.113213.
URL https://www.sciencedirect.com/science/article/pii/S0301421522004323

[9] D. A. Tejada-Arango, G. Morales-Espana, S. Wogrin, E. Centeno, Power-Based Generation Expansion Planning for Flexibility Requirements, IEEE Transactions on Power Systems 35 (3) (2020) 2012–2023. doi:10.1109/TPWRS.2019.2940286.
URL https://ieeexplore.ieee.org/document/8869877/

[10] L. Maulén, M. Castro, Lorca, M. Negrete-Pincetic, Optimization-based expansion planning for power and hydrogen systems with feedback from a unit commitment model, Applied Energy 343 (2023) 121207. doi:10.1016/j.apenergy.2023.121207.
URL https://www.sciencedirect.com/science/article/pii/S0306261923005718

[11] Y. Zhou, Q. Zhai, W. Yuan, J. Wu, Capacity expansion planning for wind power and energy storage considering hourly robust transmission constrained unit commitment, Applied Energy 302 (2021) 117570. doi:10.1016/j.apenergy.2021.117570.
URL https://www.sciencedirect.com/science/article/pii/S030626192100948X

[12] O. H. Abdalla, M. A. Abu Adma, A. S. Ahmed, Generation expansion planning considering unit commitment constraints and data-driven robust optimization under uncertainties, International Transactions on Electrical Energy Systems 31 (6) (2021) e12878, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2050-7038.12878. doi:10.1002/2050-7038.12878.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2050-7038.12878

[13] N. Helistö, J. Kiviluoma, G. Morales-España, C. O'Dwyer, Impact of operational details and temporal representations on investment planning in energy systems dominated by wind and solar, Applied Energy 290 (2021) 116712. doi:10.1016/j.apenergy.2021.116712.
URL https://www.sciencedirect.com/science/article/pii/S0306261921002312

[14] M. Hoffmann, Temporal Aggregation Methods for Energy System Modeling, Ph.D. thesis (2023). doi:10.18154/RWTH-2023-01465.
URL https://publications.rwth-aachen.de/record/945181/files/945181.pdf

[15] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renewable Energy 117 (2018) 474–487. doi:10.1016/j.renene.2017.10.017.
URL https://www.sciencedirect.com/science/article/pii/S0960148117309783

[16] K. Poncelet, E. Delarue, J. Duerinck, D. Six, W. D'haeseleer, The Importance of Integrating the Variability of Renewables in Long-term Energy Planning Models (2014).
URL https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wp-importance.pdf

[17] H. Harb, C. Schwager, R. Streblow, D. Müller, Optimal Design of Energy Systems in Residential Districts with Interconnected Local Heating And Electrical Networks, 2015. doi:10.26868/25222708.2015.2137.
URL https://publications.ibpsa.org/conference/paper/?id=bs2015_2137

[18] L. Middelhauve, N. Ljubic, J. Granacher, L. Girardin, F. Maréchal, Data reduction for mixed integer linear programming in complex energy systems, in: 34th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS 2021), ECOS 2021 Program Organizers, Taormina, Italy, 2022, pp. 267–277. doi:10.52202/062738-0024.
URL http://www.proceedings.com/062738-0024.html

[19] H. Teichgraeber, A. R. Brandt, Systematic Comparison of Aggregation Methods for Input Data Time Series Aggregation of Energy Systems Optimization Problems, in: M. R. Eden, M. G. Ierapetritou, G. P. Towler (Eds.), Computer Aided Chemical Engineering, Vol. 44 of 13 International Symposium on Process Systems Engineering (PSE 2018), Elsevier, 2018, pp. 955–960. doi:10.1016/B978-0-444-64241-7.50154-3.
URL https://www.sciencedirect.com/science/article/pii/B9780444642417501543

[20] P. M. Barnaghi, A. Abu Bakar, Z. A. Othman, Enhanced symbolic aggregate approximation method for financial time series data representation, in: 2012 6th International Conference on New Trends in Information Science, Service Science and Data Mining (ISSDM2012), 2012, pp. 790–795.
URL https://ieeexplore.ieee.org/document/6528740/?arnumber=6528740

[21] C. Guo, H. Li, D. Pan, An Improved Piecewise Aggregate Approximation Based on Statistical Features for Time Series Mining, in: Y. Bi, M.-A. Williams (Eds.), Knowledge Science, Engineering and Management, Springer, Berlin, Heidelberg, 2010, pp. 234–244. doi:10.1007/978-3-642-15280-1_23.

[22] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: a novel symbolic representation of time series, Data Mining and Knowledge Discovery 15 (2) (2007) 107–144. doi:10.1007/s10618-007-0064-z.
URL http://link.springer.com/10.1007/s10618-007-0064-z

[23] B. Lkhagva, Y. Suzuki, K. Kawagoe, Extended SAX: extension of symbolic aggregate approximation for financial time series data representation, Proceeding of IEICE the 17th Data Engineering Workshop (2006).

[24] S. Pfenninger, Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability, Applied Energy 197 (2017) 1–13. doi:10.1016/j.apenergy.2017.03.051.
URL https://www.sciencedirect.com/science/article/pii/S0306261917302775

[25] H. Ren, X. Liao, Z. Li, A. AI-Ahmari, Anomaly detection using piecewise aggregate approximation in the amplitude domain, Applied Intelligence 48 (5) (2018) 1097–1110. doi:10.1007/s10489-017-1017-x.
URL http://link.springer.com/10.1007/s10489-017-1017-x

[26] S. Kirschbaum, M. Powilleit, M. Schotte, F. Özbeg, Efficient Solving of Time-Coupled Energy System MILP Models Using a Problem Specific LP Relaxation, in: 36th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS 2023), ECOS 2023, Las Palmas De Gran Canaria, Spain, 2023, pp. 2774–2785. doi:10.52202/069564-0249.
URL http://www.proceedings.com/069564-0249.html

[27] R. Yokoyama, Y. Shinano, Y. Wakayama, T. Wakui, Model reduction by time aggregation for optimal design of energy supply systems by an MILP hierarchical branch and bound method, Energy 181 (2019) 782–792. doi:10.1016/j.energy.2019.04.066.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360544219306899

[28] R. Yokoyama, Y. Shinano, S. Taniguchi, M. Ohkura, T. Wakui, Optimization of energy supply systems by MILP branch and bound method in consideration of hierarchical relationship between design and operation, Energy Conversion and Management 92 (2015) 92–104. doi:10.1016/j.enconman.2014.12.020.
URL https://www.sciencedirect.com/science/article/pii/S0196890414010498

[29] B. Lim, S. Zohren, Time-series forecasting with deep learning: a survey, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 379 (2194) (2021) 20200209, publisher: Royal Society. doi:10.1098/rsta.2020.0209.
URL https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0209

[30] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, J. C. Riquelme, Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting, Applied Sciences 10 (7) (2020) 2322, number: 7 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/app10072322.
URL https://www.mdpi.com/2076-3417/10/7/2322

[31] A. Asfaw, B. Simane, A. Hassen, A. Bantider, Variability and time series trend analysis of rainfall and temperature in northcentral Ethiopia: A case study in Woleka sub-basin, Weather and Climate Extremes 19 (2018) 29–41. doi:10.1016/j.wace.2017.12.002.
URL https://www.sciencedirect.com/science/article/pii/S2212094717300932

[32] M. Mudelsee, Trend analysis of climate time series: A review of methods, Earth-Science Reviews 190 (2019) 310–322. doi:10.1016/j.earscirev.2018.12.005.
URL https://www.sciencedirect.com/science/article/pii/S0012825218303726

[33] J.-R. Jiang, J.-B. Kao, Y.-L. Li, Semi-Supervised Time Series Anomaly Detection Based on Statistics and Deep Learning, Applied Sciences 11 (15) (2021) 6698, number: 15 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/app11156698.
URL https://www.mdpi.com/2076-3417/11/15/6698

[34] H. Teichgraeber, A. R. Brandt, Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities, Renewable and Sustainable Energy Reviews 157 (2022) 111984. doi:10.1016/j.rser.2021.111984.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032121012478

[35] O. Omoyele, M. Hoffmann, M. Koivisto, M. Larrañeta, J. M. Weinand, J. Linßen, D. Stolten, Increasing the resolution of solar and wind time series for energy system modeling: A review, Renewable and Sustainable Energy Reviews 189 (2024) 113792. doi:10.1016/j.rser.2023.113792.
URL https://www.sciencedirect.com/science/article/pii/S1364032123006494

[36] G. Morales-Espana, A. Ramos, J. Garcia-Gonzalez, An MIP Formulation for Joint Market-Clearing of Energy and Reserves Based on Ramp Scheduling, IEEE Transactions on Power Systems 29 (1) (2014) 476–488. doi:10.1109/TPWRS.2013.2259601.
URL https://ieeexplore.ieee.org/document/6514884/

[37] G. Morales-España, C. Gentile, A. Ramos, Tight MIP formulations of the power-based unit commitment problem, OR Spectrum 37 (4) (2015) 929–950. doi:10.1007/s00291-015-0400-4.
URL http://link.springer.com/10.1007/s00291-015-0400-4

[38] F. Neumann, T. Brown, The near-optimal feasible space of a renewable power system model, Electric Power Systems Research 190 (2021) 106690. doi:10.1016/j.epsr.2020.106690.
URL https://linkinghub.elsevier.com/retrieve/pii/S0378779620304934

[39] L. Göke, M. Kendziorski, Adequacy of time-series reduction for renewable energy systems, Energy 238 (2022). doi:10.1016/j.energy.2021.121701.

[40] O. Omoyele, S. Matrone, M. Hoffmann, E. Ogliari, J. M. Weinand, S. Leva, D. Stolten, Impact of temporal resolution on the design and reliability of residential energy systems, Energy and Buildings 319 (2024) 114411. doi:10.1016/j.enbuild.2024.114411.
URL https://www.sciencedirect.com/science/article/pii/S0378778824005279

[41] J. Wu, P. B. Luh, Y. Chen, M. A. Bragin, B. Yan, A Novel Optimization Approach for Sub-Hourly Unit Commitment With Large Numbers of Units and Virtual Transactions, IEEE Transactions on Power Systems 37 (5) (2022) 3716–3725, conference Name: IEEE Transactions on Power Systems. doi:10.1109/TPWRS.2021.3137842.
URL https://ieeexplore.ieee.org/abstract/document/9662252

[42] B. Shirizadeh, P. Quirion, Do multi-sector energy system optimization models need hourly temporal resolution? A case study with an investment and dispatch model applied to France, Applied Energy 305 (2022) 117951. doi:10.1016/j.apenergy.2021.117951.
URL https://www.sciencedirect.com/science/article/pii/S0306261921012617

[43] D. Liu, X. Li, P.-C. Huang, Y. Gu, Y. Ling, K. Zhong, R. Liu, X. Chen, L. Liang, Downsizing Without Downgrading: Approximated Dynamic Time Warping on Nonvolatile Memories, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 39 (1) (2020) 131–144. doi:10.1109/TCAD.2018.2878182.
URL https://ieeexplore.ieee.org/document/8515082/

[44] V. Mladenov, V. Chobanov, A. Georgiev, Impact of Renewable Energy Sources on Power System Flexibility Requirements, Energies 14 (10) (2021) 2813, number: 10 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/en14102813.
URL https://www.mdpi.com/1996-1073/14/10/2813

[45] C.-C. Chia, Z. Syed, Using Adaptive Downsampling to Compare Time Series with Warping, in: 2010 IEEE International Conference on Data Mining Workshops, IEEE, Sydney, TBD, Australia, 2010, pp. 1304–1311. doi:10.1109/ICDMW.2010.94.
URL http://ieeexplore.ieee.org/document/5693444/

[46] K. R. Shahapure, C. Nicholas, Cluster Quality Analysis Using Silhouette Score, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 747–748. doi:10.1109/DSAA49011.2020.00096.
URL https://ieeexplore.ieee.org/abstract/document/9260048

[47] Planning the future grid - TYNDP.
URL https://tyndp.entsoe.eu/

[48] Technology catalogues (Aug. 2016).
URL https://ens.dk/en/our-services/technology-catalogues

[49] Mathematical Formulation · TulipaEnergyModel.jl.
URL https://TulipaEnergy.github.io/TulipaEnergyModel.jl/mathematical-formulation/

Appendix A. Algorithms Down-Sampling

Below the algorithms used for the general down-sampling procedures in the basic statistics and hybrid methods are displayed.

Algorithm 1: General Down-Sampling Procedure

```
Function downsample_time_series:
  data, N, strategy
Input  : data: hourly time series {(t_i, y_i)}_{i=1}^M
         N: resampling interval in hours
         strategy: down-sampling strategy
Output: result: down-sampled time series with M time steps

result ← initialize_series_with_same_indices_as(data) ;
for i range(0, M, N) do
    block ← data[i:i+N] ;
    resampled_value ← block.strategy() ;
    result[i:i+N-1] ← resampled_value ;
result ← fill_missing_data(result) ;
return result ;
```

Algorithm 2: Hybrid Down-Sampling Procedure

```
Function downsample_time_series_hybrid:
  data, N, std_threshold
Input  : data: hourly time series {(t_i, y_i)}_{i=1}^M
         N: resampling interval in hours
         std_threshold: standard deviation threshold
Output: result: down-sampled time series with M time steps

result ← initialize_series_with_same_indices_as(data) ;
for i range(0, M, N) do
    block ← data[i:i+N] ;
    std_block ← calculate_std(block) ;
    if std_block < std_threshold then
        resampled_value ← benchmark(block) ;
    else
        resampled_value ← max(block) ;
    result[i:i+N-1] ← resampled_value ;
result ← fill_missing_data(result) ;
return result ;
```

Appendix B. Tulipa Version 0.7 Mathematical Formulation[4]

The tables below summarise the sets, parameters and variables used by the Tulipa tool in version 0.7. Additionally, the mathematical formulation used in the Tulipa tool is summarised. Please note that the Tables and mathematical formulation are all reproduced from [49].

Appendix B.1. ***Overview of Sets, Parameters and Variables***

[4] https://tulipaenergy.github.io/TulipaEnergyModel.jl/v0.7/mathematical-formulation/

Table B.4: Overview of all sets used in the Tulipa formulation

Name	Description	Elements
A	Energy assets	$a \in A$
A_c	Consumer energy assets	$A_c \subseteq A$
A_p	Producer energy assets	$A_p \subseteq A$
A_s	Storage energy assets	$A_s \subseteq A$
A_h	Hub energy assets (e.g., transshipment)	$A_h \subseteq A$
A_{cv}	Conversion energy assets	$A_{cv} \subseteq A$
A_i	Energy assets with investment method	$A_i \subseteq A$
F	Flow connections between two assets	$f \in F$
F_t	Transport flow between two assets	$F_t \subseteq F$
F_i	Transport flow with investment method	$F_i \subseteq F_t$
$F_{in}(a)$	Set of flows going into asset a	$F_{in}(a) \subseteq F$
$F_{out}(a)$	Set of flows going out of asset a	$F_{out}(a) \subseteq F$
RP	Representative periods	$rp \in RP$
K	Time steps within rp	$k \in K$

Table B.5: Overview of all parameters used in the Tulipa formulation

Name	Domain	Description
$p_a^{\text{investment\_cost}}$	A_i	Investment cost of asset units [kEUR/MW/year]
$p_a^{\text{investment\_limit}}$	A_i	Investment limit of asset units [MW]
$p_a^{\text{unit\_capacity}}$	A	Capacity of asset units [MW]
$p_a^{\text{peak\_demand}}$	A_c	Peak demand [MW]
$p_a^{\text{init\_capacity}}$	A	Initial capacity of asset units [MW]
$p_f^{\text{investment\_cost}}$	F_i	Investment cost of flow connections [kEUR/MW/year]
$p_f^{\text{variable\_cost}}$	F	Variable cost of flow connections [kEUR/MWh]
$p_f^{\text{unit\_capacity}}$	F_t	Capacity increment for flow connections [MW]
$p_f^{\text{init\_export\_capacity}}$	F_t	Initial export capacity of flow connections [MW]
$p_f^{\text{init\_import\_capacity}}$	F_t	Initial import capacity of flow connections [MW]
p_{rp}^{weight}	RP	Representative period weight [h]
$p_{a,rp,k}^{\text{profile}}$	A, RP, K	Asset profile [p.u.]
$p_{f,rp,k}^{\text{profile}}$	F, RP, K	Flow connections profile [p.u.]
$p_a^{\text{ene\_to\_pow\_ratio}}$	A_s	Energy to power ratio [h]
$p_a^{\text{init\_storage\_level}}$	A_s	Initial storage level [MWh]
p_a^{inflow}	A_s	Energy storage inflows [MWh]
p_f^{eff}	F	Flow efficiency [p.u.]

Table B.6: Overview of all variables used in the Tulipa formulation

Name	Domain	Description
$v_{f,rp,k}^{\text{flow}}$	F, RP, K	Flow between two assets [MW]
$v_a^{\text{investment}}$	A_i	Number of installed asset units [units]
$v_f^{\text{investment}}$	F_i	Number of installed units between two assets [units]
$s_{a,rp,k}^{\text{level}}$	A_s, RP, K	Storage level [MWh]

Appendix B.2. ***Objective Function***

$$\text{minimize} \quad \text{assets\_investment\_cost} + \text{flows\_investment\_cost} + \text{flows\_variable\_cost} \tag{B.1}$$

$$\text{assets\_investment\_cost} = \sum_{a \in A} \text{investment\_cost}_a \cdot \text{unit\_capacity}_a \cdot v_a^{\text{investment}} \tag{B.2}$$

$$\text{flows\_investment\_cost} = \sum_{f \in F} \text{investment\_cost}_f \cdot \text{unit\_capacity}_f \cdot v_f^{\text{investment}} \tag{B.3}$$

$$\text{flows\_variable\_cost} = \sum_{f \in F} \sum_{rp \in RP} \sum_{k \in K} \text{weight}_{rp} \cdot \text{variable\_cost}_f \cdot v^{\text{flow}}_{f,rp,k} \tag{B.4}$$

Appendix B.3. ***Balancing Constraints for Consumer Energy Assets***

$$\sum_{f \in F^{\text{in}}(a)} v^{\text{flow}}_{f,rp,k} - \sum_{f \in F^{\text{out}}(a)} v^{\text{flow}}_{f,rp,k} = \text{profile}_{a,rp,k} \cdot \text{peak\_demand}_a, \quad \forall a \in A_c, \forall rp \in RP, \forall k \in K \tag{B.5}$$

Appendix B.4. ***Balancing Constraints for Storage Energy Assets***

$$\text{level}_{a,rp,k} = \text{level}_{a,rp,k-1} + \text{inflow}_{a,rp,k} + \sum_{f \in F^{\text{in}}(a)} \text{eff}_f \cdot v^{\text{flow}}_{f,rp,k} - \sum_{f \in F^{\text{out}}(a)} \frac{1}{\text{eff}_f} \cdot v^{\text{flow}}_{f,rp,k}, \quad \forall a \in A_s, \forall rp \in RP, \forall k \in K \tag{B.6}$$

Appendix B.5. ***Balancing Constraints for Hub Energy Assets***

$$\sum_{f \in F^{\text{in}}(a)} v^{\text{flow}}_{f,rp,k} = \sum_{f \in F^{\text{out}}(a)} v^{\text{flow}}_{f,rp,k}, \quad \forall a \in A_h, \forall rp \in RP, \forall k \in K \tag{B.7}$$

Appendix B.6. ***Balancing Constraints for Conversion Energy Assets***

$$\sum_{f \in F^{\text{in}}(a)} \text{eff}_f \cdot v^{\text{flow}}_{f,rp,k} = \sum_{f \in F^{\text{out}}(a)} v^{\text{flow}}_{f,rp,k} \cdot \text{eff}_f, \quad \forall a \in A_{cv}, \forall rp \in RP, \forall k \in K \tag{B.8}$$

Appendix B.7. ***Capacity Constraints: Maximum Output Flows Limit***

$$\sum_{f \in F^{\text{out}}(a)} v^{\text{flow}}_{f,rp,k} \leq \text{profile}_{a,rp,k} \cdot \left(\text{init\_capacity}_a + \text{unit\_capacity}_a \cdot v^{\text{investment}}_a\right), \quad \forall a \in A_{cv} \cup A_s \cup A_p, \forall rp \in RP, \forall k \in K \tag{B.9}$$

Appendix B.8. ***Capacity Constraints: Maximum Input Flows Limit***

$$\sum_{f \in F^{\text{in}}(a)} v^{\text{flow}}_{f,rp,k} \leq \text{profile}_{a,rp,k} \cdot \left(\text{init\_capacity}_a + \text{unit\_capacity}_a \cdot v^{\text{investment}}_a\right), \quad \forall a \in A_s, \forall rp \in RP, \forall k \in K \tag{B.10}$$

Appendix B.9. ***Capacity Constraints: Lower Bound on Flows***

$$v^{\text{flow}}_{f,rp,k} \geq 0, \quad \forall f \notin F_t, \forall rp \in RP, \forall k \in K \tag{B.11}$$

Appendix B.10. ***Transport Flow Capacity: Maximum Transport Flow Limit***

$$v^{\text{flow}}_{f,rp,k} \leq \text{profile}_{f,rp,k} \cdot \left(\text{init\_export\_capacity}_f + \text{unit\_capacity}_f \cdot v^{\text{investment}}_f\right), \quad \forall f \in F_t, \forall rp \in RP, \forall k \in K \tag{B.12}$$

Appendix B.11. ***Transport Flow Capacity: Minimum Transport Flow Limit***

$$v^{\text{flow}}_{f,rp,k} \geq -\text{profile}_{f,rp,k} \cdot \left(\text{init\_import\_capacity}_f + \text{unit\_capacity}_f \cdot v^{\text{investment}}_f\right), \quad \forall f \in F_t, \forall rp \in RP, \forall k \in K \tag{B.13}$$

Appendix B.12. ***Extra Constraints for Energy Storage Assets: Maximum Storage Level Limit***

$$0 \leq \text{level}_{a,rp,k} \leq \text{init\_storage\_capacity}_a + \text{ene\_to\_pow\_ratio}_a \cdot \text{unit\_capacity}_a \cdot v^{\text{investment}}_a, \quad \forall a \in A_s, \forall rp \in RP, \forall k \in K \tag{B.14}$$

Appendix B.13. ***Extra Constraints for Energy Storage Assets: Cycling Constraints***

$$\text{level}_{a,rp,k=K} \geq \text{init\_storage\_level}_a, \quad \forall a \in A_s, \forall rp \in RP \tag{B.15}$$

Appendix B.14. ***Investment Constraints: Maximum Investment Limit for Assets***

$$v_a^{\text{investment}} \leq \text{investment\_limit}_a, \quad \forall a \in A_i \tag{B.16}$$

Appendix B.15. ***Investment Constraints: Maximum Investment Limit for Flows***

$$v_f^{\text{investment}} \leq \text{investment\_limit}_f, \quad \forall f \in F_i \tag{B.17}$$

Appendix C. Input Data NWEU Case Study

The table below summarises a description of all input files that are needed to set up a case study with version 0.7 of the Tulipa[5].

Table C.7: Overview of all input CSV files required to build a case study with TulipaEnergyModel tool

File Name	Contents & Purpose
assets-data.csv	Lists all assets in the system with their characteristics. Includes information on the asset's name, type, whether it is active or investable, investment and operational costs, capacity, and specific parameters for storage assets such as initial storage capacity and energy -to-power ratio.
flows-data.csv	Defines the flows between assets, detailing the type of carrier, source, and destination assets, if the flow is active or investable, costs, capacities, and efficiencies. It also includes connection flows between countries. This file illustrates the interactions and energy exchanges within the system.
assets-rep-periods -profiles.csv	Contains information on profiles for each asset. Specifies the asset, profile type and profile name.
flows-rep-periods-profiles.csv	Contains information on input profiles for each flow. Specifies the asset, profile type and profile name.
assets-rep-periods-partitions.csv	Describes how to divide the representative period into time blocks for assets, with options for uniform, explicit, or mathematical partitioning. This file allows users to specify the temporal resolution for modelling each asset.
flows-rep-periods-partitions.csv	Similar to assets-partitions.csv but focuses on the partitioning of time for flows between assets.

[5] https://tulipaenergy.github.io/TulipaEnergyModel.jl/v0.7/how-to-use/

rep-periods-data.csv	Outlines the representative periods used in the model, including their ID, number of time steps, and resolution. This file sets the framework for how time is represented in the model, allowing for the aggregation of similar periods into representative blocks.
rep-periods-mapping.csv	Maps the full time periods to the representative periods, providing weights for how representative each period is. This file is crucial for understanding the significance of each representative period within the broader temporal context of the model.
profiles-rep-periods -availability.csv	Includes the profiles of the availability of renewable energy sources in p.u. value. The profiles that are included here are the availability of solar, wind onshore and offshore in each country included in the model. For each profile type and each country 8760 time steps are included (thus hourly). Depending on the resolution set in the partitions files, e.g. 2,3 or 4 hourly resolution is considered for the profiles.
profiles-rep-periods -demand.csv	Includes the profiles of the demand in p.u. value, for each country included in the model. For each profile 8760 time steps are included, per year. Depending on the resolution set in the partitions files, e.g. 2,3 or 4 hourly resolution is considered for the profiles.
profiles-rep-periods -inflows.csv	Includes the profiles of the hydro reservoir, hydro run of river and hydro open pumped storage in p.u. value, for each country included in the model. For each profile 8760 time steps are included (thus hourly). Depending on the resolution set in the partitions files, e.g. 2,3 or 4 hourly resolution is considered for the profiles.

Appendix D. List of Basic Statistics Experiments

The table below provides an overview of all down-sampling experiments that were performed using the basic statistics methods. All input profiles were down-sampled to 2, 3 and 4-hourly resolution and used in the operational model configuration (op), the investment model configuration (inv) and the increased limits investment configuration (inv_increased).

Table D.8: List of experiments performed with benchmark and basic statistics down-sampling

Name	**Down-Sampling**	**Resolutions**	**Profiles**	**Models**
Benchmark	average	2,3,4	all	op, inv, inv_increased
Basic statistics	first	2,3,4	all	op, inv, inv_increased
	last	2,3,4	all	op, inv, inv_increased
	median	2,3,4	all	op, inv, inv_increased
	maximum	2,3,4	all	op, inv, inv_increased
	minimum	2,3,4	all	op, inv, inv_increased
	linear interpolation	2,3,4	all	op, inv, inv_increased
	midpoint	2,3,4	all	op, inv, inv_increased

Appendix E. Threshold Range and GridSearch Hybrid Down-Sampling

The ranges of the possible threshold are shown below for the Dutch profiles. Note that the determined thresholds are based on the Dutch profiles, but are used also for the profiles specific to other countries. At the lower end of the range, the minimum value of the threshold ensures that the hybrid method will take the maximum in each block and thus will be the same as the basic statistics max methods. This is similar at the high end of the range, where the hybrid method will pick the benchmark in every block and therefore will be the same as the benchmark benchmark method.

- Energy Demand: (=max) 0 < threshold < 0.045 (=benchmark)
- Solar: (=max) 0 < threshold < 0.15 (=benchmark)
- Wind Onshore: (=max) 0 < threshold < 0.18 (=benchmark)
- Wind Offshore: (=max) 0 < threshold < 0.25 (=benchmark)

The threshold range was explored by taking 4 equally distanced values within the range for each profile. For solar this is 0.03, 0.06, 0.09 and 0.12. For the energy demand this is 0.01, 0.02, 0.03, 0.04. For onshore wind this was 0.03, 0.06, 0.09 and 0.12. Lastly, for wind offshore this was 0.05, 0.1, 0.15 and 0.2. In addition, a more advanced exploration of the threshold range was performed for each profile in each resolution using grid search and solving for CAQE of the ramps (ramp representation), see Equation 12. Figure E.25 and E.26 display the results of the grid search. The thresholds that were used from the grid search are specific by profile and resolution, see below.

- Energy Demand: 2h threshold 0.014, 3h threshold 0.2, 4h threshold 0.17
- Solar: 2h threshold 0.025, 3h threshold 0.05, 4h threshold 0.08
- Wind Onshore: 2h threshold 0.4, 3h threshold 0.375, 4h threshold 0.35
- Wind Offshore: 2h threshold 0.01, 3h threshold 0.015, 4h threshold 0.025

Below the results of the grid search described above are presented, showing for which thresholds the CAQE ramp is at minimum.

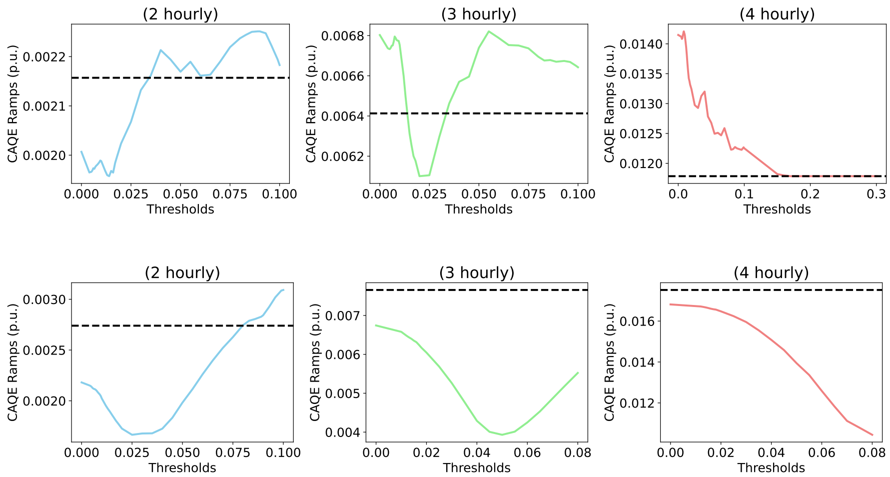


Figure E.25: Results of the grid search for the thresholds associated with the optimal CAQE ramps for the energy demand and solar profiles.

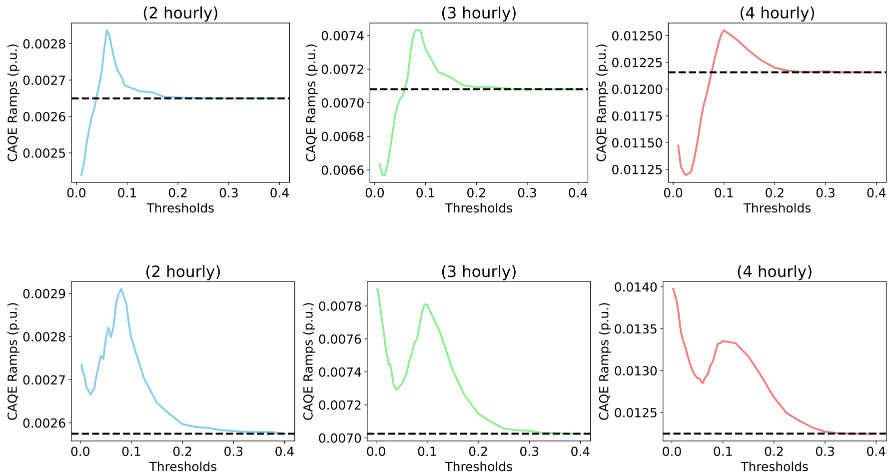


Figure E.26: Results of the grid search for the thresholds associated with the optimal CAQE ramps for the wind onshore and wind offshore profiles.

Appendix F. List of Hybrid Experiments

The table below summarises all experiments that were performed with the hybrid method. In each experiment, only 1 profile was down-sampled to separate the effect of down-sampling on specific profiles. The thresholds that were used in each experiment are listed, and each profile was down-sampled to 2, 3 and 4 hourly resolution. Note that the optimal thresholds are specific by resolution. The experiments were performed in the operational model configuration (op), the investment model configuration (inv) and the increased limits investment configuration (inv_increased).

Table F.9: List of hybrid down-sampling experiments

Profile	**Thresholds**	**Other Profiles Downsampled**	**Resolution**	**Models**	**sample name**
energy demand	0.01	benchmark	2,3,4	op, inv, inv_increased	ed_001
energy demand	0.014 (optimal grid search)	benchmark	2	op, inv, inv_increased	ed_0014
energy demand	0.02 (optimal grid search 3h res)	benchmark	2,3,4	op, inv, inv_increased	ed_002
energy demand	0.03	benchmark	2,3,4	op, inv, inv_increased	ed_003
energy demand	0.04	benchmark	2,3,4	op, inv, inv_increased	ed_004
energy demand	0.17 (optimal grid search)	benchmark	4	op, inv, inv_increased	ed_017
solar	0.025 (optimal grid search)	benchmark	2	op, inv, inv_increased	s_0025
solar	0.03	benchmark	2,3,4	op, inv, inv_increased	s_003
solar	0.05 (optimal grid search)	benchmark	3	op, inv, inv_increased	s_005
solar	0.06	benchmark	2,3,4	op, inv, inv_increased	s_006
solar	0.08 (optimal grid search)	benchmark	4	op, inv, inv_increased	s_008
solar	0.09	benchmark	2,3,4	op, inv, inv_increased	s_009
solar	0.12	benchmark	2,3,4	op, inv, inv_increased	s_012

wind offshore	0.05	benchmark	2,3,4	op, inv, inv_increased	woff_005
wind offshore	0.1	benchmark	2,3,4	op, inv, inv_increased	woff_01
wind offshore	0.15	benchmark	2,3,4	op, inv, inv_increased	woff_015
wind offshore	0.2	benchmark	2,3,4	op, inv, inv_increased	woff_02
wind	0.35 (optimal grid search)	benchmark	4	op, inv, inv_increased	woff_035
wind offshore	0.375 (optimal grid search)	benchmark	3	op, inv, inv_increased	woff_0375
wind offshore	0.4 (optimal grid search)	benchmark	2	op, inv, inv_increased	woff_04
wind onshore	0.01 (optimal grid search)	benchmark	2	op, inv, inv_increased	woff_001
wind onshore	0.03	benchmark	2,3,4	op, inv, inv_increased	won_003
wind onshore	0.06	benchmark	2,3,4	op, inv, inv_increased	won_006
wind onshore	0.09	benchmark	2,3,4	op, inv, inv_increased	won_009
wind onshore	0.12	benchmark	2,3,4	op, inv, inv_increased	won_012

Appendix G. Thresholds Hybrid Experiments

The table below summarises the thresholds used for each profile on each resolution, the thresholds are ordered from low to high and labelled 1 to 5 in the same way as they are labelled in the results matrices for clarity.

Table G.10: Thresholds used for hybrid experiments

profile	threshold 1	threshold 2	threshold 3	threshold 4	threshold 5
energy demand (2h)	0.01	0.014	0.02	0.03	0.04
solar (2h)	0.025	0.03	0.06	0.09	0.12
wind onshore (2h)	0.01	0.03	0.06	0.09	0.12
wind offshore (2h)	0.05	0.2	0.1	0.15	0.4
energy demand (3h)	0.01	0.02	0.03	0.04	0.17
solar (3h)	0.03	0.05	0.06	0.09	0.12
wind onshore (3h)	0.015	0.03	0.06	0.09	0.12
wind offshore (3h)	0.05	0.2	0.1	0.15	0.375
energy demand (4h)	0.01	0.02	0.02	0.03	0.04
solar (4h)	0.03	0.05	0.06	0.08	0.09
wind onshore (4h)	0.025	0.03	0.06	0.09	0.12
wind offshore (4h)	0.05	0.2	0.1	0.15	0.35

Appendix H. Visual Overview Methods Quality Check

The figure below gives a schematic overview of the methods used to perform the quality checks in the investment and operational model configurations. In the investment configuration, the investment decisions are extracted from the down-sampled model and fixed in a model on hourly resolution. The operational decisions were then evaluated based on the energy not served. In the operational configuration model, the operational decisions of the down0sampled model were extracted and used as input in an hourly model, limiting the availability for the generation assets. The operational decisions were evaluated using the energy not served.

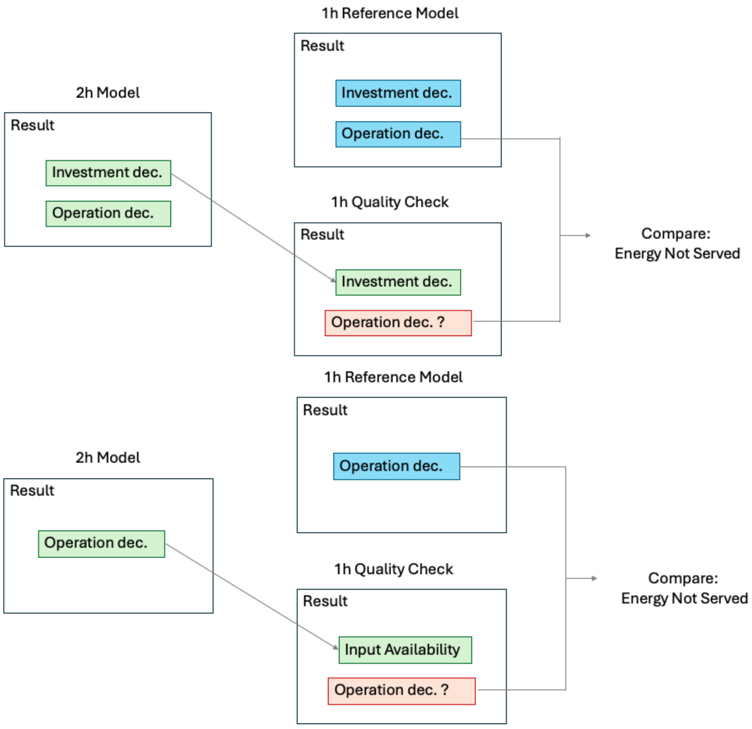


Figure H.27: Schematic overview of the methods for the quality check for the investment models (top) and operation models (bottom).

Appendix I. Results of Euclidean Distance and Clustering Analysis

The table below summarises the Euclidean distances from the year 2008 for climate years from 1982-2016 in p.u.. The years 1997 and 1991 are most different from 2008, however after performing k-means clustering with an optimal cluster number of three the most different year from 2008 in cluster 3 was used, since 1991 was in the same cluster as 1997, see Figure I.28.

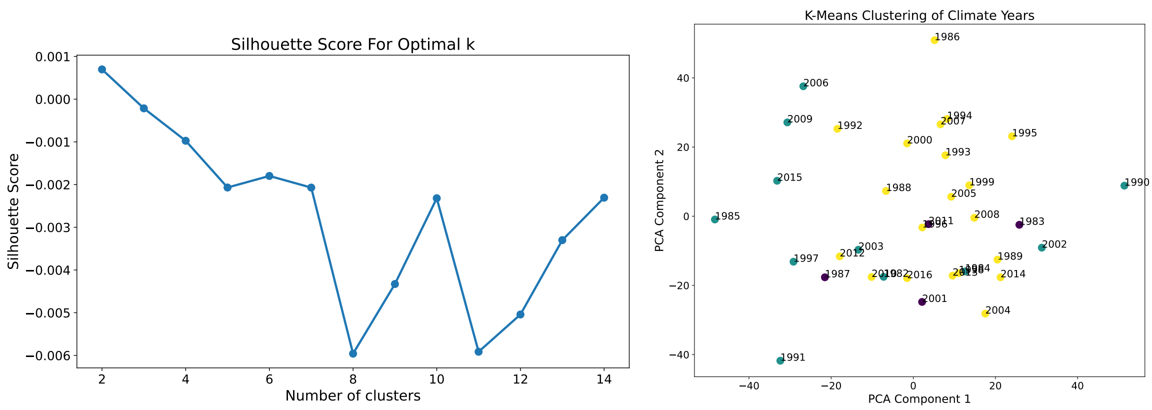


Figure I.28: The silhouette score (left) of the k-means clustering, shows that the optimal cluster number is 2 (highest score) (left). The visualisation of k-means clustering with PCA (right), shows that the reference year 2008 and the two most different years (1887 and 1991) using Euclidean distance are in the same clusters. Thus the most different year from 2008 in terms of Euclidean distance in the third cluster was picked, which is 2011 (right).

Table I.11: Euclidean distances of all climate years from 1982 to 2016 compared to 2008

Year	Euclidean Distance (p.u.)
1997	150.691573
1991	146.592648
1996	146.061514
1986	145.441309
2012	145.376498
1985	145.191372
1994	144.856760
1988	144.029480
2000	143.802728
2015	143.705783
2011	143.570686
2006	143.552203
1990	143.276239
2007	143.229413
1987	143.206293
1989	143.055902
2010	142.687578
2004	142.079269
2003	142.033186
2013	141.743327
1993	141.593094
2009	141.491615
1982	140.975944
2001	140.570150
1999	140.015032
1998	140.003114
2002	139.990847
1992	139.552008
2014	138.657603
1983	138.332700
1995	137.614374
2005	137.413259
2016	136.862339
1984	136.718304

Appendix J. Initial Representation Analysis

In the subsections below a visual overview of the locations of minima and maxima in the down-sampled energy demand, solar, and wind profiles are shown, including the hybrid and basic statistics experiments. Moreover for the wind profiles the generation and ramp distribution curves are presented.

Appendix J.1. ***Energy Demand Profile***

Appendix J.1.1. ***Minima and Maxima***

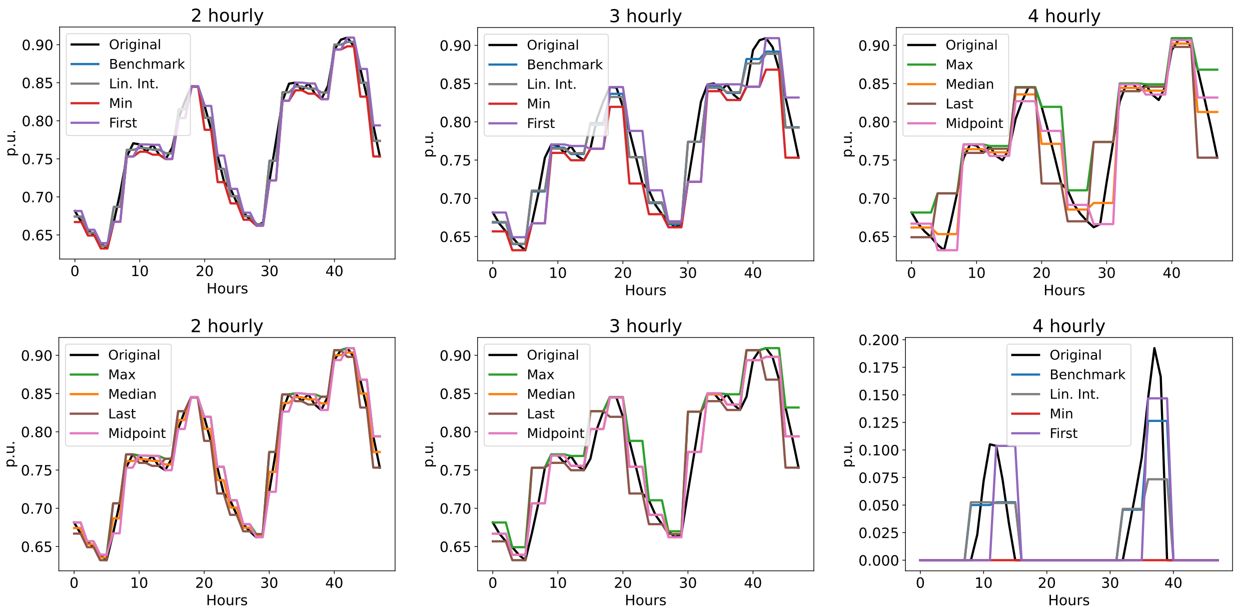


Figure J.29: Locations of minima and maxima in the energy demand profile compared to the original Dutch energy demand profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution.

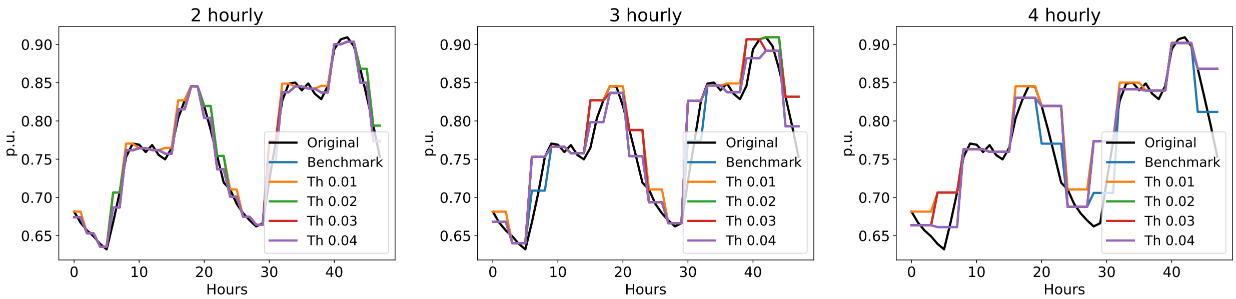


Figure J.30: Locations of minima and maxima in the energy demand profile compared to the original Dutch energy demand profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution.

Appendix J.2. ***Solar Profile***

Appendix J.2.1. ***Minima and Maxima***

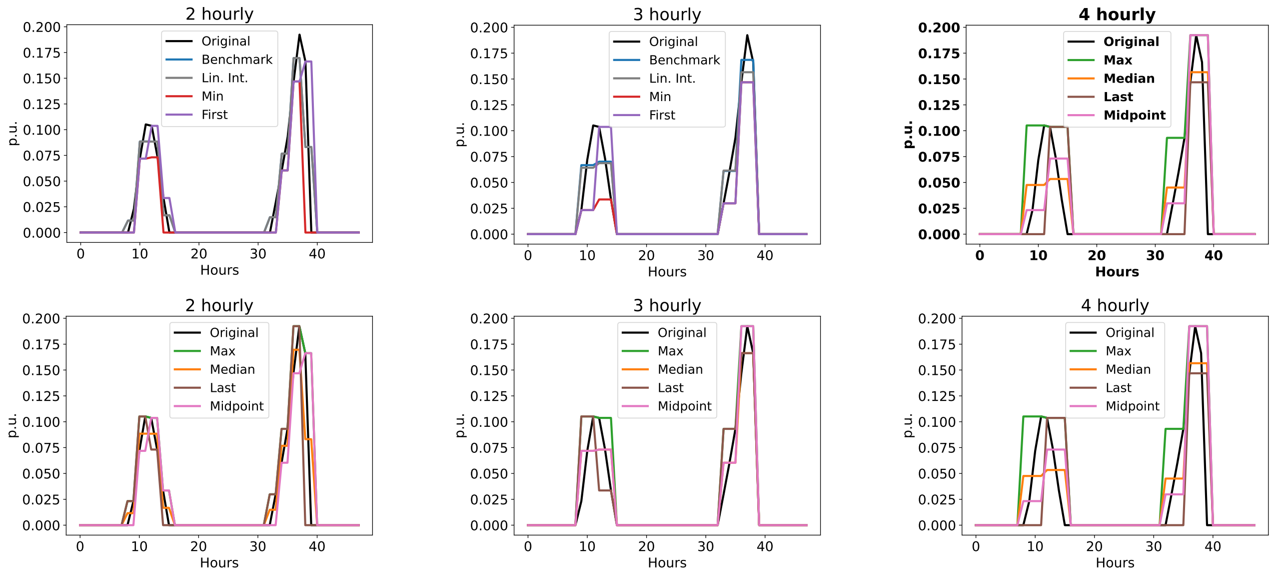


Figure J.31: Locations of minima and maxima in the solar profile compared to the original Dutch solar profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution.

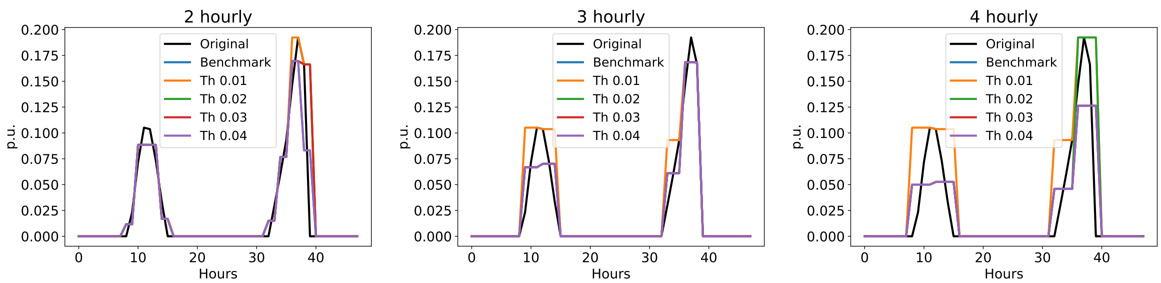


Figure J.32: Locations of minima and maxima in the solar profile compared to the original Dutch solar profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution.

Appendix J.3.1. ***Minima and Maxima***

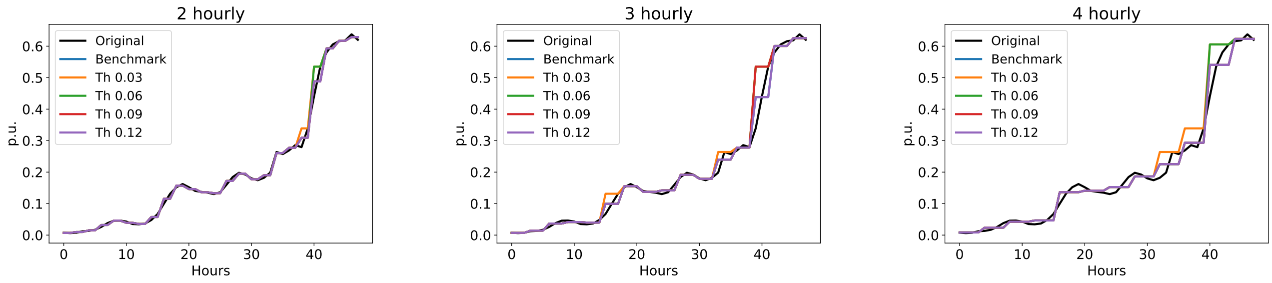


Figure J.33: Locations of minima and maxima in the wind onshore profile compared to the original Dutch wind onshore profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution.

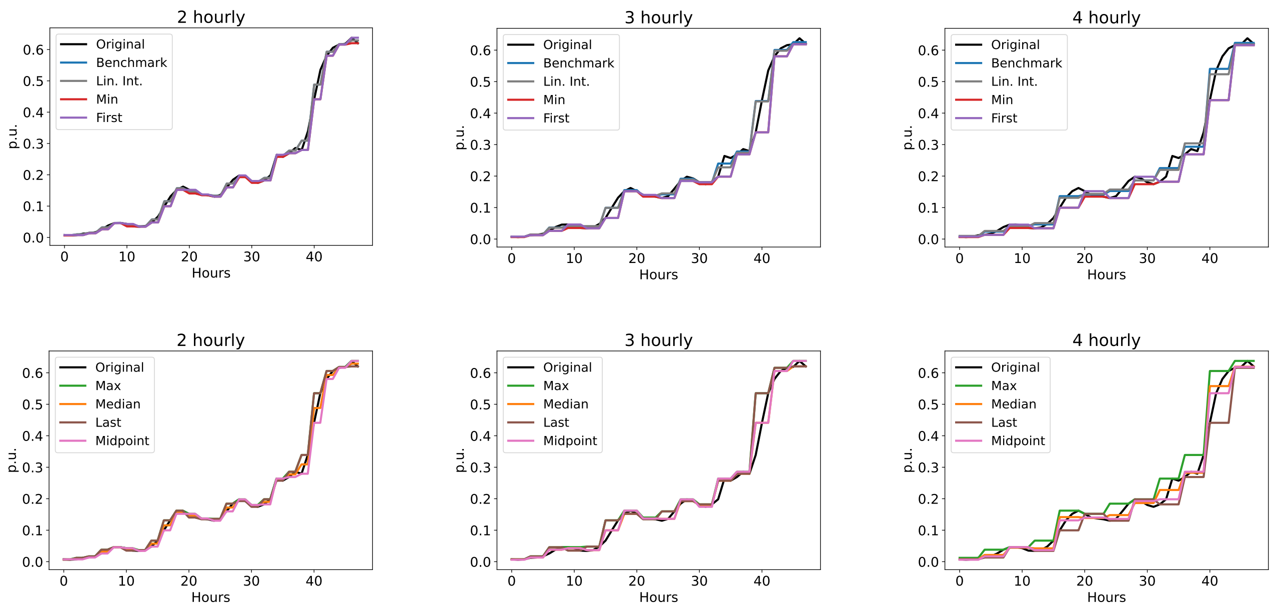


Figure J.34: Locations of minima and maxima in the wind onshore profile compared to the original Dutch wind onshore profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution.

Appendix J.3.2. *Generation and Ramp Duration Curves*

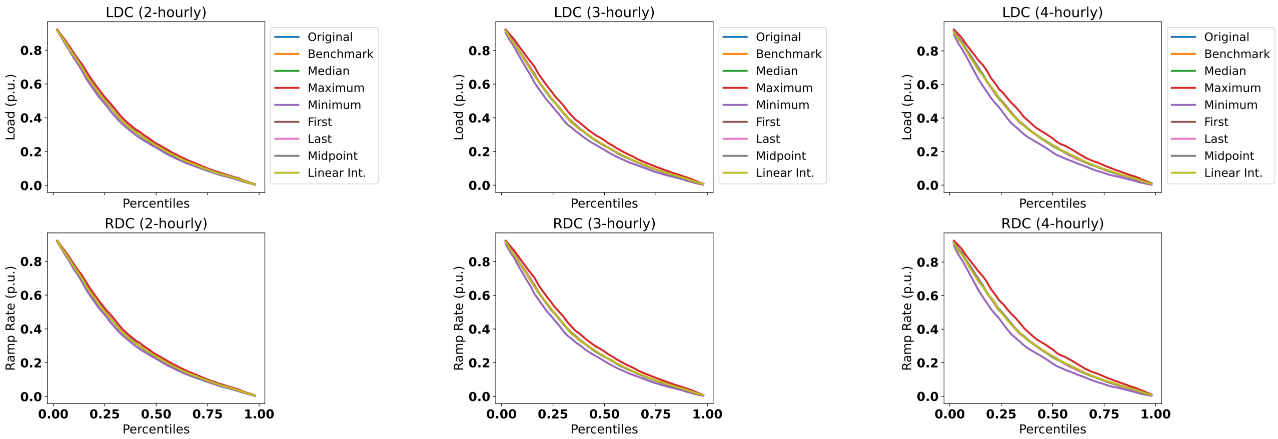


Figure J.35: The generation and ramp duration curves of wind onshore profiles down-sampled with basic statistics methods to 2, 3 and 4-hourly resolution.

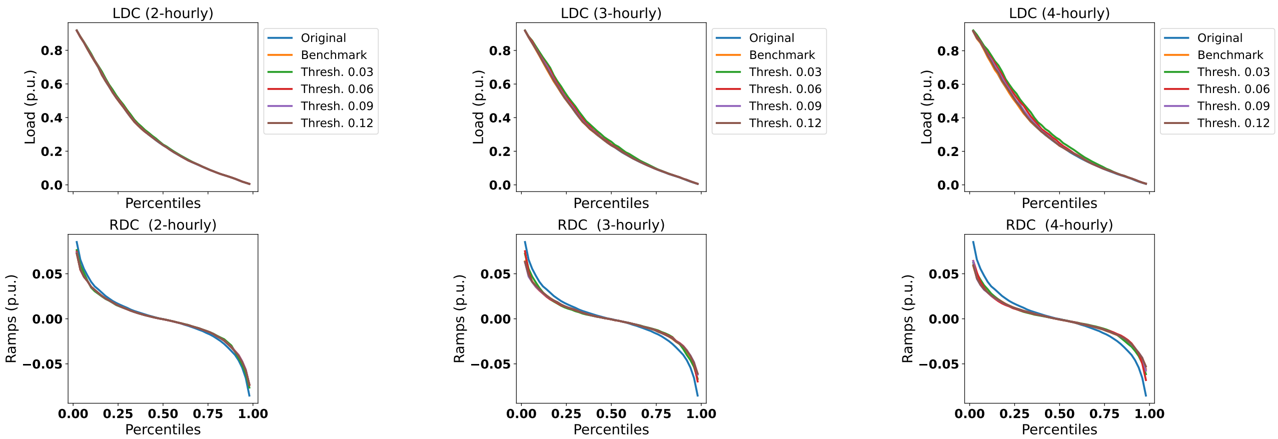


Figure J.36: The generation and ramp duration curves of wind onshore profiles down-sampled with the hybrid method to 2, 3 and 4-hourly resolution.

Appendix J.4. ***Wind Offshore Profile***

Appendix J.4.1. ***Minima and Maxima***

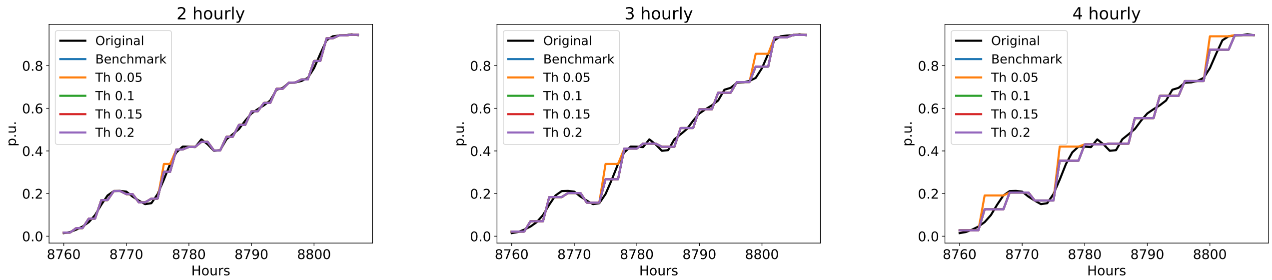


Figure J.37: Locations of minima and maxima in the wind offshore profile compared to the original Dutch wind offshore profile. The profiles were down-sampled with the hybrid method, to 2, 3 and 4 hourly resolution.

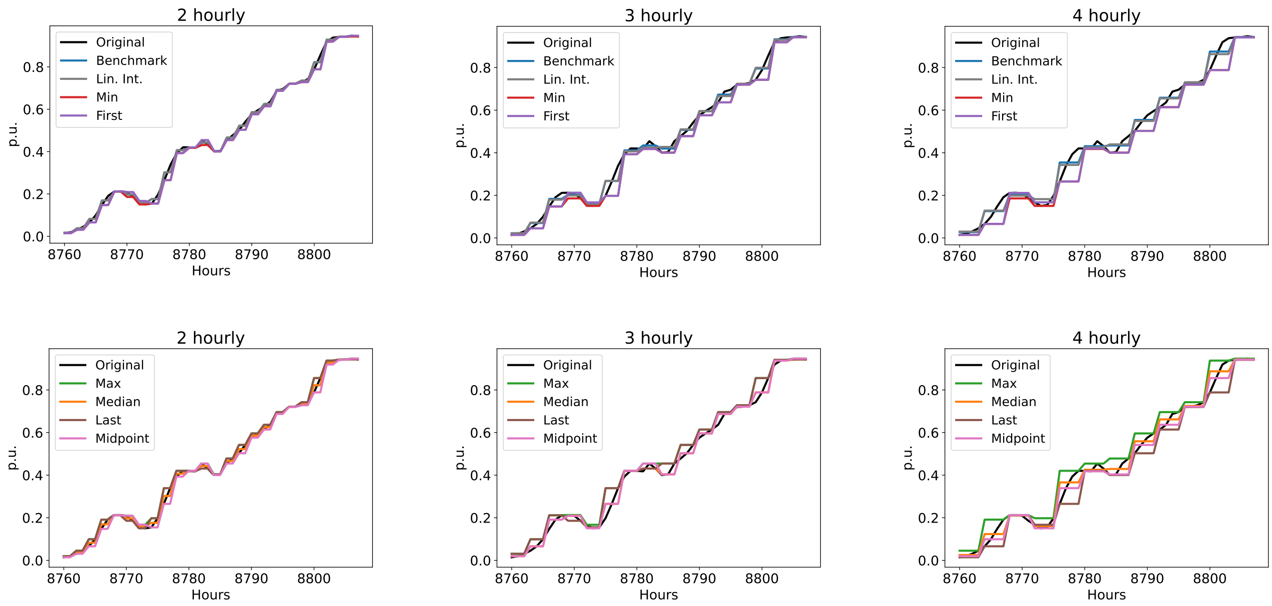


Figure J.38: Locations of minima and maxima in the wind offshore profile compared to the original Dutch wind offshore profile. The profiles were down-sampled with the basic statistics methods, to 2, 3 and 4 hourly resolution.

Appendix J.4.2. ***Generation and Ramp Duration Curves***

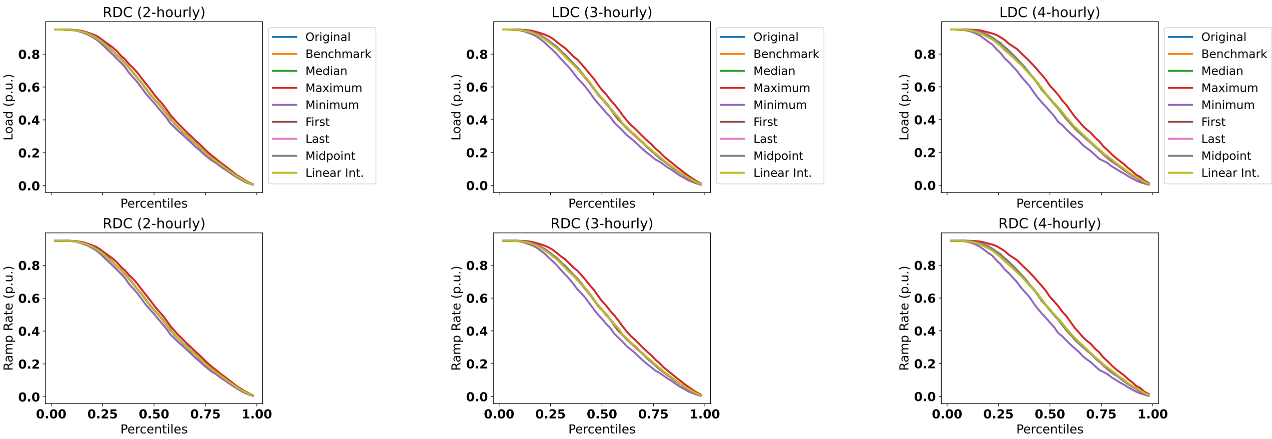


Figure J.39: The generation and ramp duration curves of wind offshore profiles down-sampled with the basic statistics methods to 2, 3 and 4-hourly resolution.

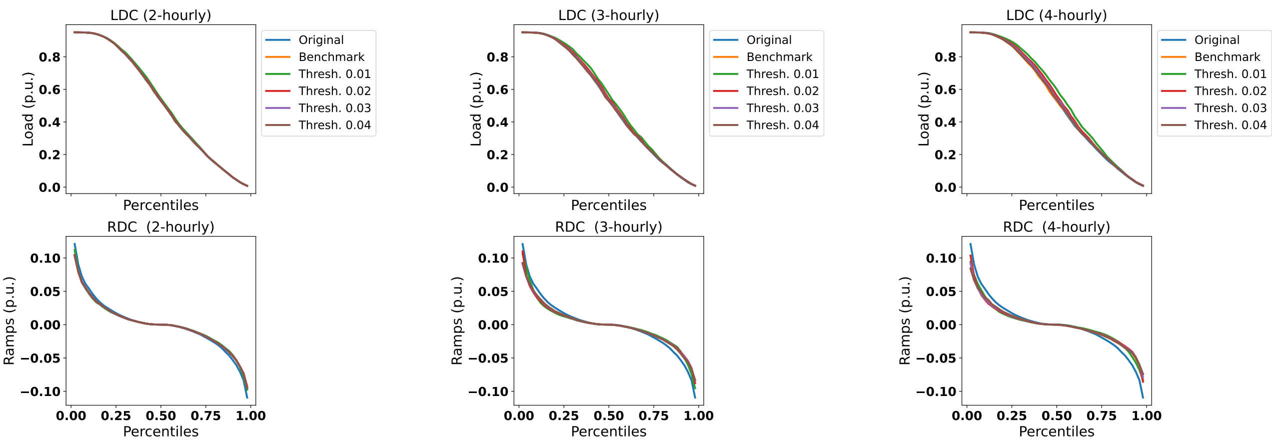


Figure J.40: The generation and ramp duration curves of wind offshore profiles down-sampled with the hybrid method to 2, 3 and 4-hourly resolution.

Appendix K. Flow Breakdown Operational Models 2008

In the matrices below the flow breakdown in different experiments is compared to the flow breakdown on hourly resolution. Cells indicated in red show that the flows are higher in the down-sampled model and flows in blue show the flows are lower in the down-sampled model. In the top panel, the comparison is made with down-sampling to 2-hourly resolution, in the middle panel on 3-hourly resolution and in the bottom panel on 4-hourly resolution. Figure K.41, the flows on hourly resolution in the operational model are compared to using the benchmark (mean/average) for aggregation. Figure K.42, the flows on hourly resolution in the operational model are compared to using the minimum for aggregation. Figure K.41, the flows on hourly resolution in the operational model are compared to using the hybrid method on the energy demand profile using a threshold of 0.04 for aggregation. Figure K.41, the flows on hourly resolution in the operational model are compared to using the hybrid method on the solar profile using a threshold of 0.06 for aggregation.

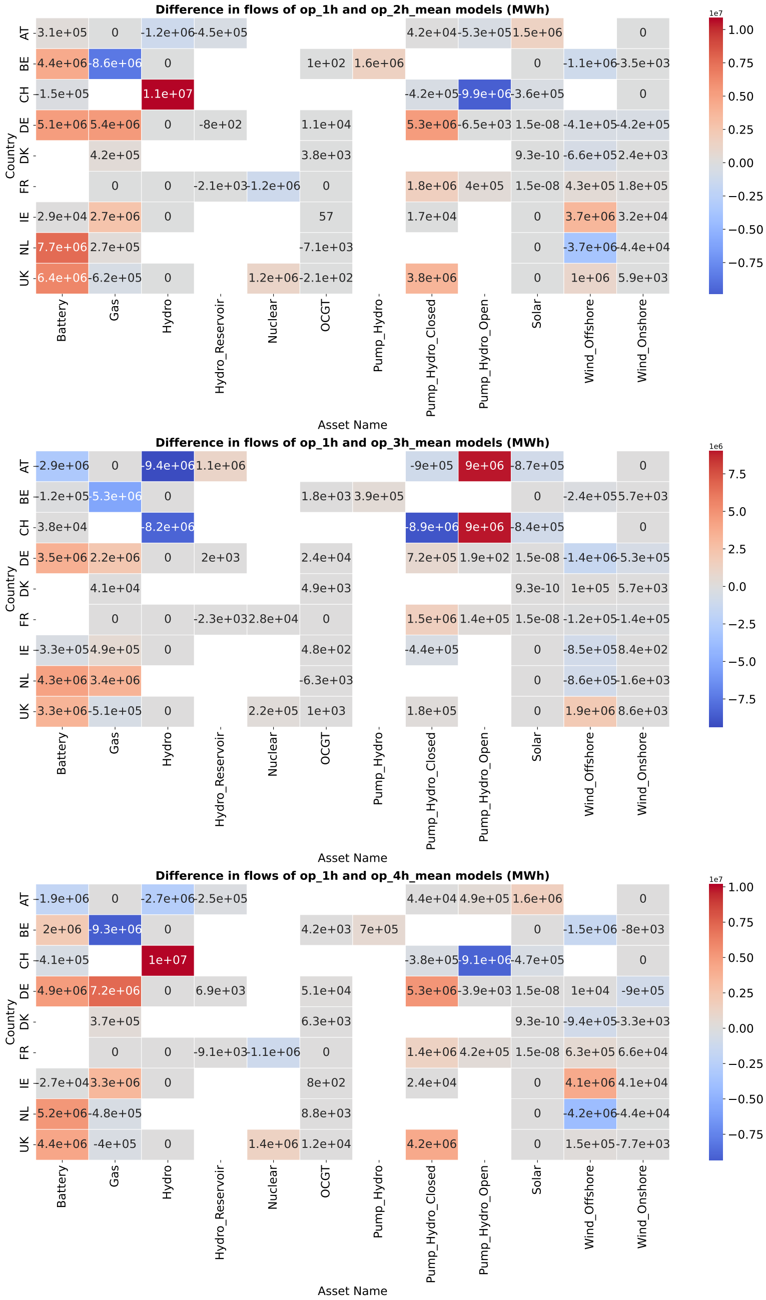


Figure K.41: A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the benchmark method on 2, 3 and 4-hourly resolution.

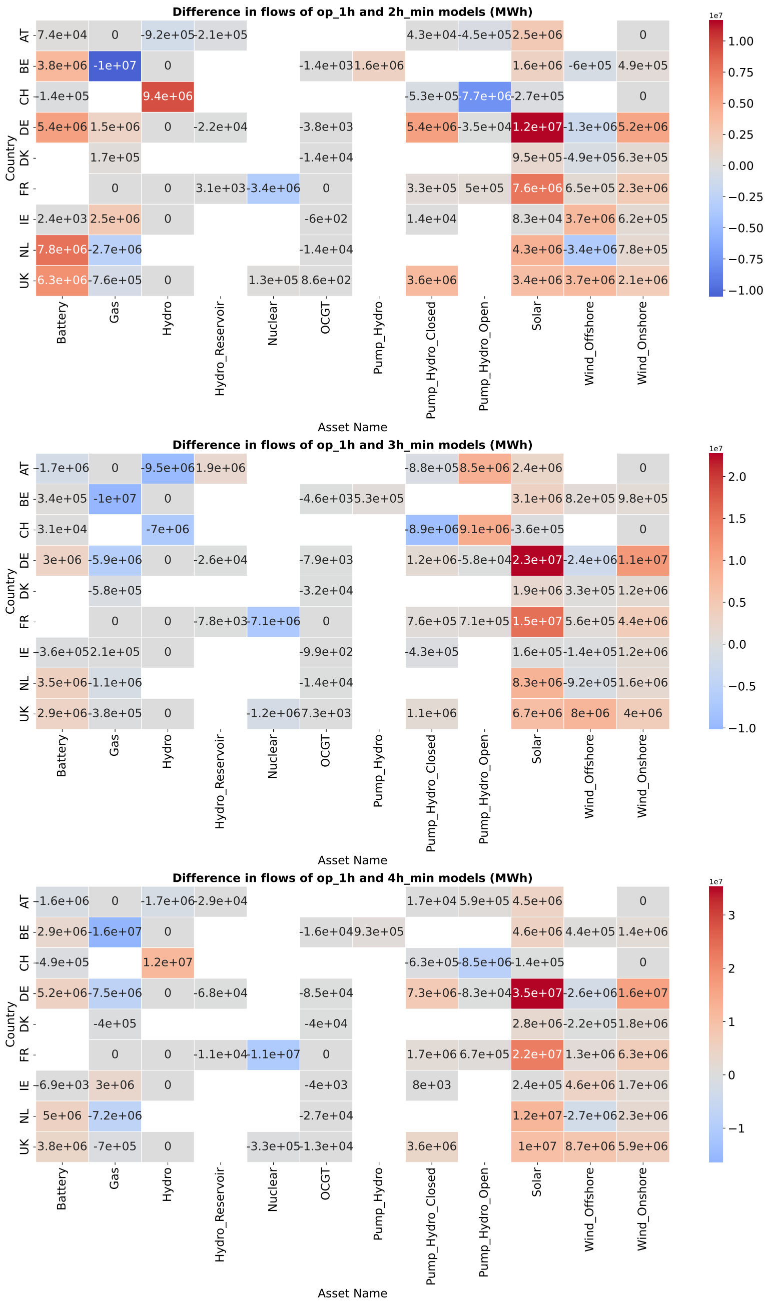


Figure K.42: A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the minimum method on 2, 3 and 4-hourly resolution.

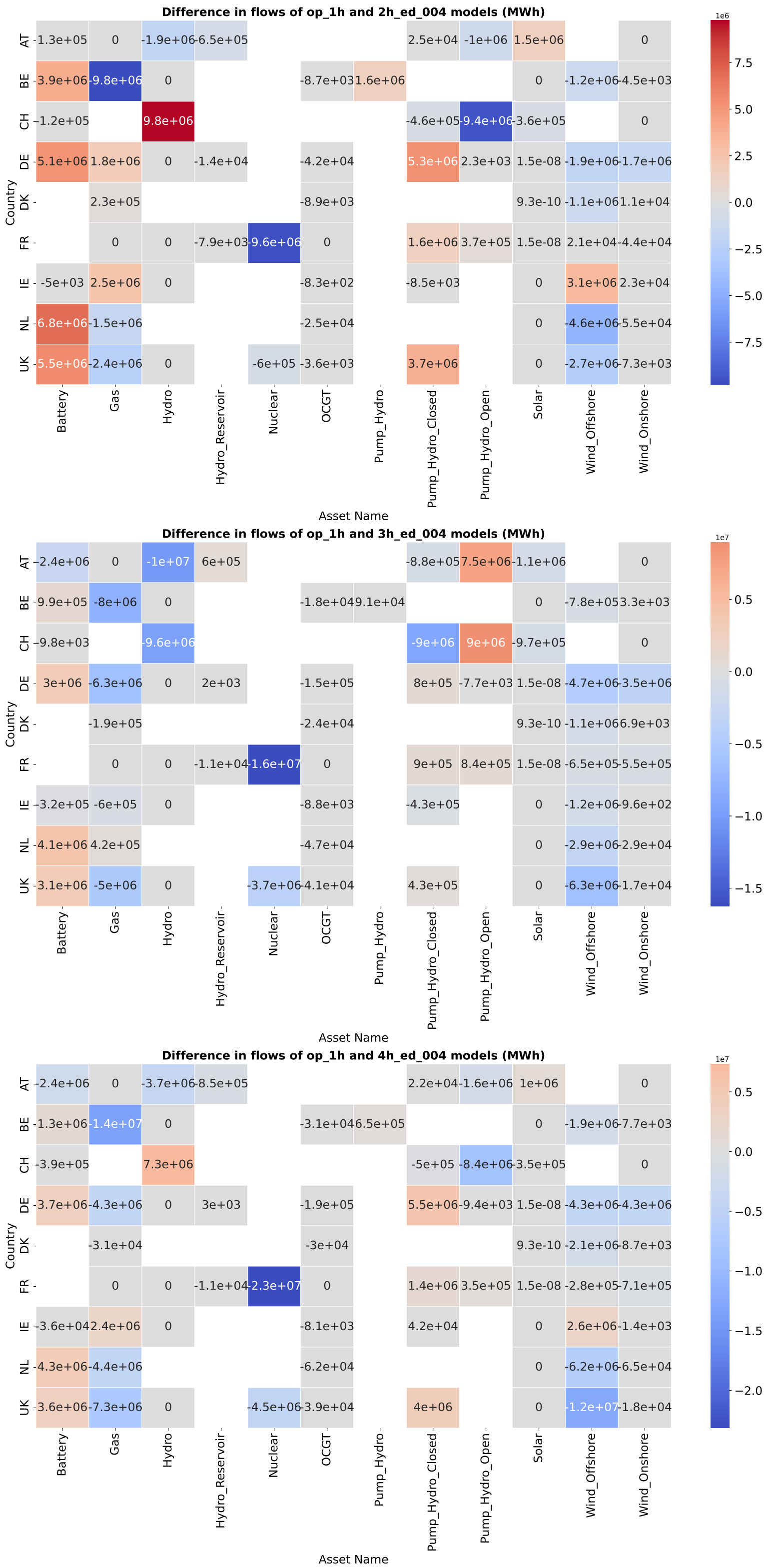


Figure K.43: A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the hybrid method on the energy demand profile with a threshold of 0.04 on 2, 3 and 4-hourly resolution.

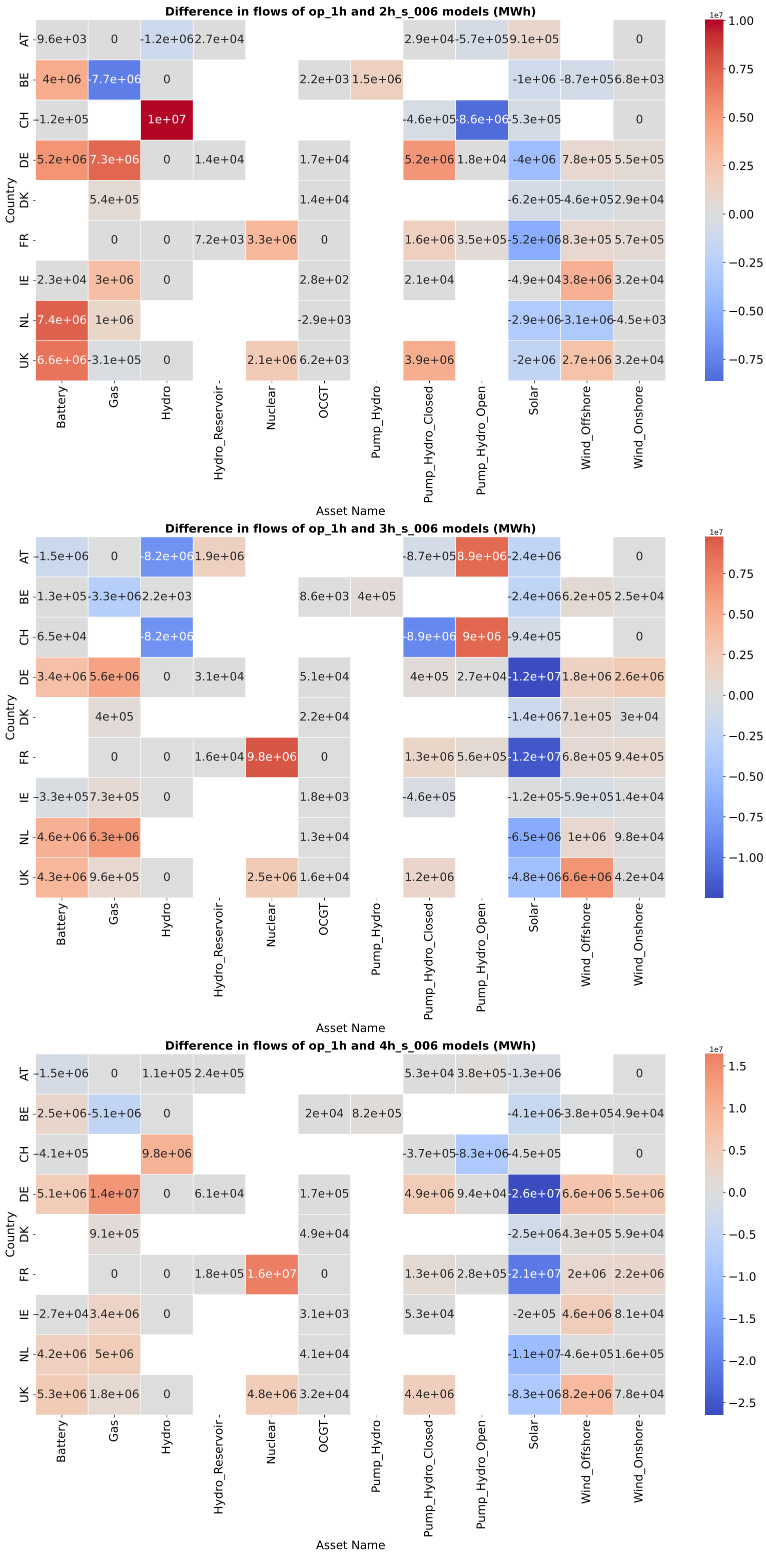


Figure K.44: A breakdown of the flow behaviour in the operational model. The flow behaviour of the hourly reference model is compared to the flow behaviour using the hybrid method on the solar profile with a threshold of 0.06 on 2, 3 and 4-hourly resolution.

Appendix L. Investment Decisions Models 2008

The figures below show the investment decisions made in the hourly reference models in different countries in the ivnestment model configurations.

Appendix L.1. ***Hourly Reference Models***

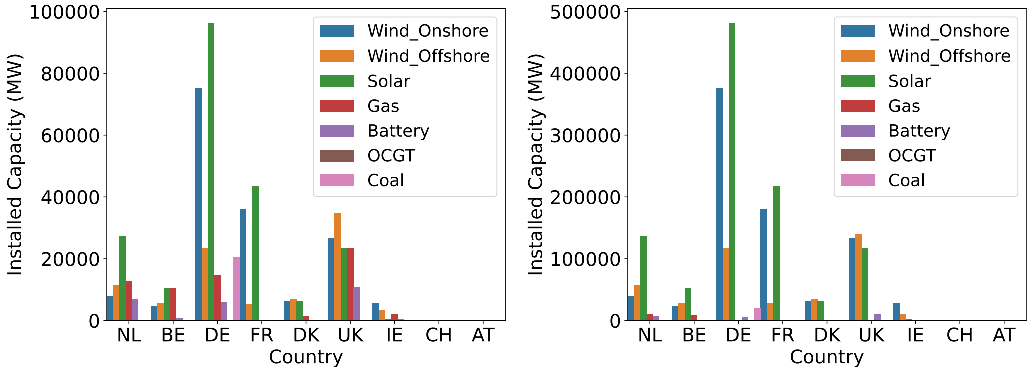


Figure L.45: Investment decisions of the hourly reference investment model (left) and of the hourly reference investment model with increased limits (right).

Appendix L.2. ***Basic Statistics***

The figures below show the investment decisions made in the basic statistics down-sampling experiments compared to the hourly reference (in blue) and to the benchmark (in orange labelled as the mean).

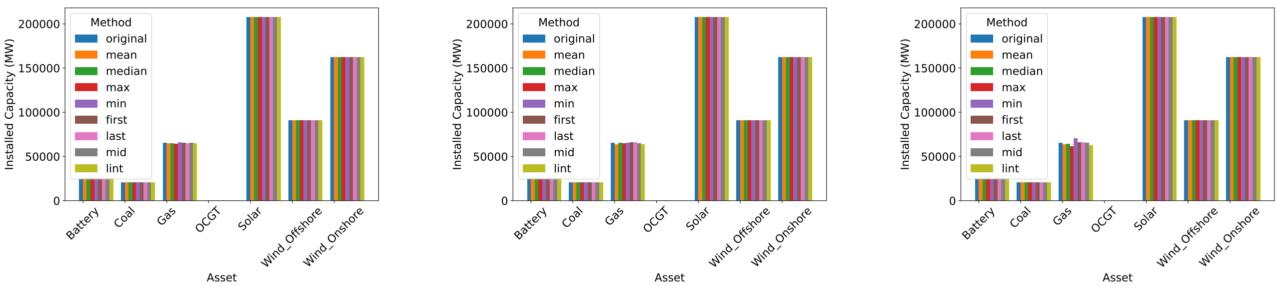


Figure L.46: Investment decisions in the investment model with basic statistics down-sampling experiments.

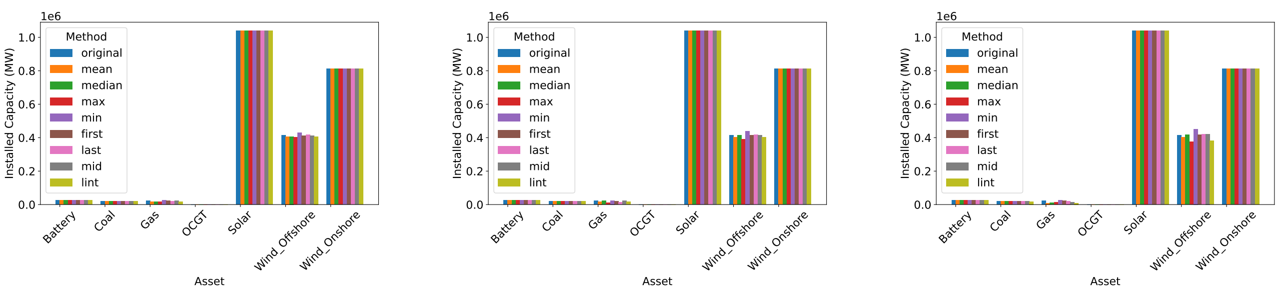


Figure L.47: Investment decisions in the increased limits investment model with basic statistics down-sampling experiments

Appendix L.3. ***Hybrid***

The figures below show the investment decisions made in the hybrid down-sampling experiments compared to the hourly reference (in blue) and to the benchmark (in black labelled as the mean).

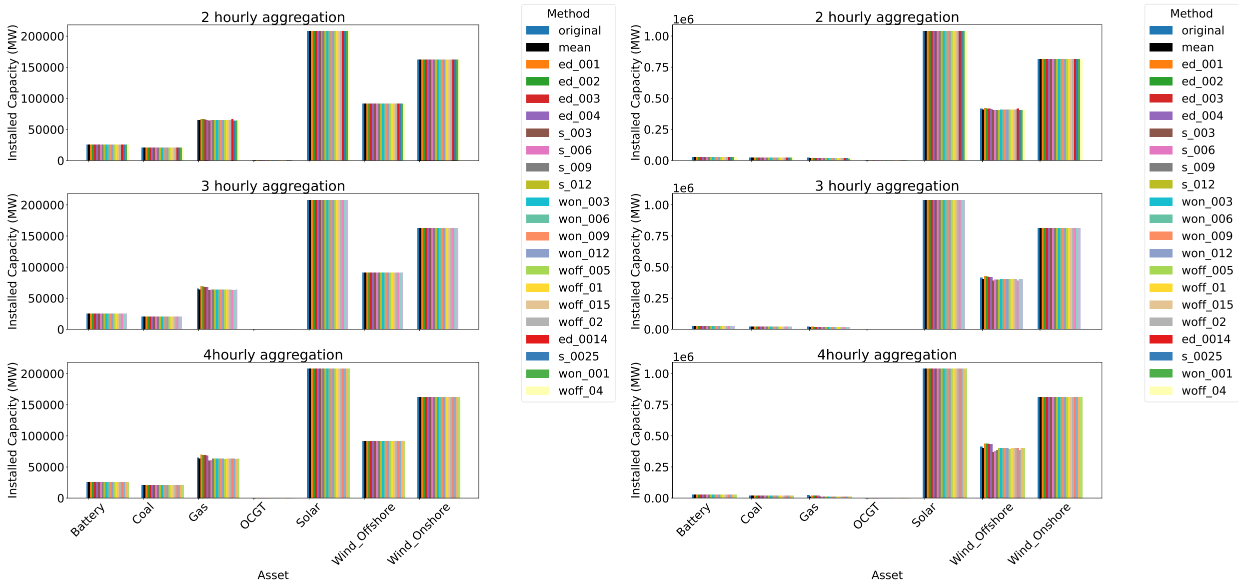


Figure L.48: Investment decisions in the investment model (left) and in the investment model with increased limits (right) with the hybrid method experiments.

Appendix L.4. ***Initial Representation Analysis 2011***

The figures below show the results of the initial representation analysis for the down-sampled 2011 input profiles. It can be observed that for both the CAQE and CAQE ramps and for both the basic statistics and hybrid methods the trends align with those observed in the initial representation analysis in 2008.

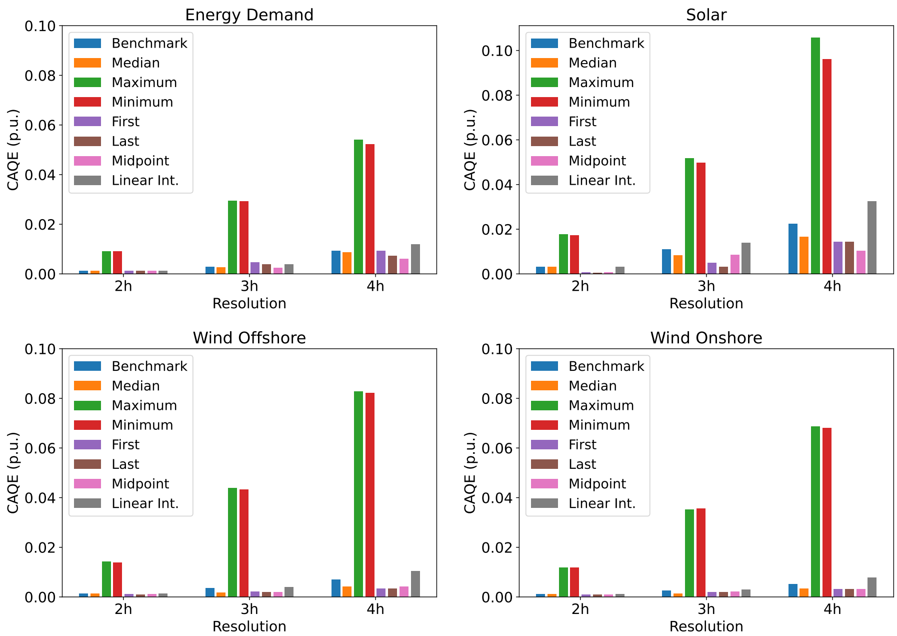


Figure L.49: Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

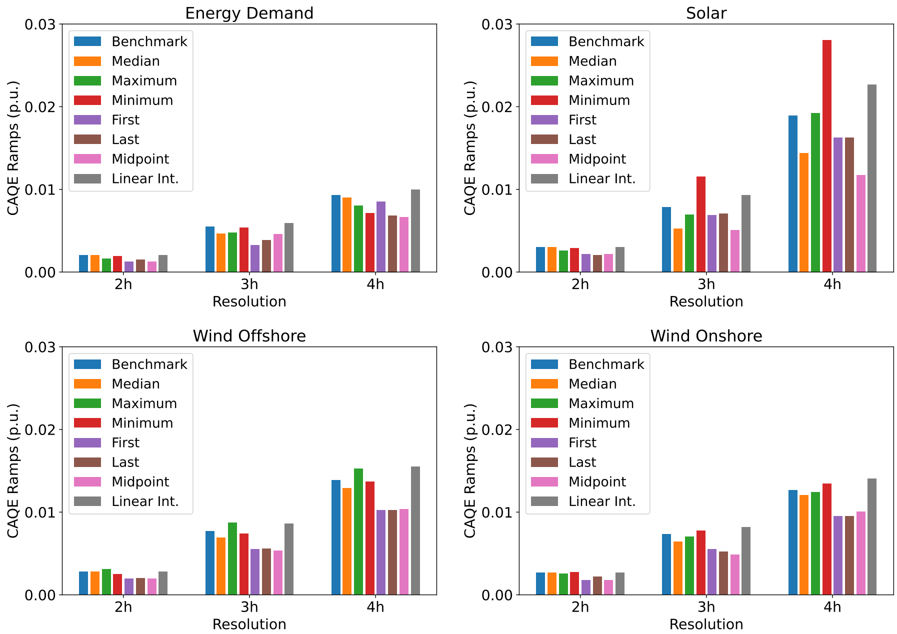


Figure L.50: Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

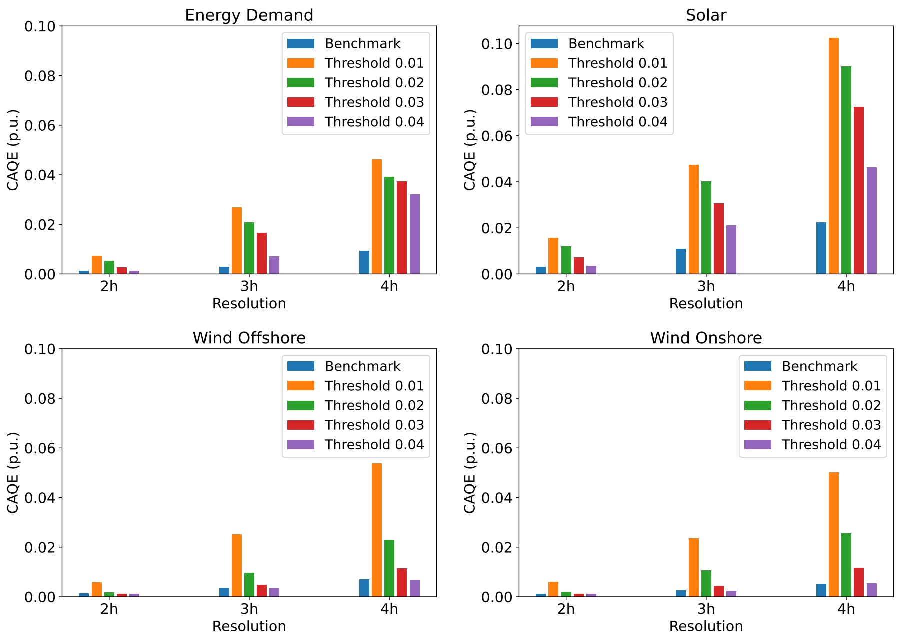


Figure L.51: Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

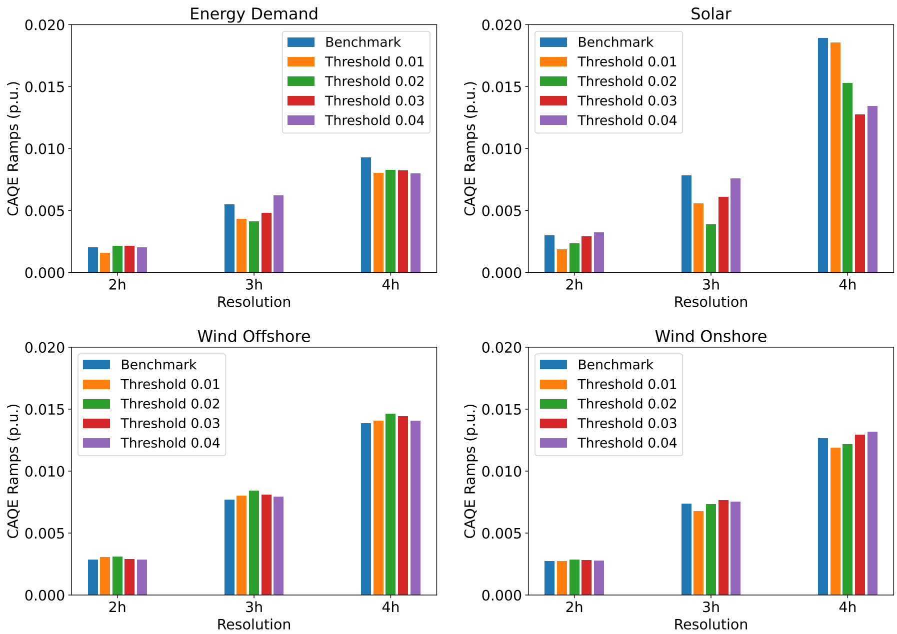


Figure L.52: Initial representation analysis results for 2011. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid method, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

The figures below show the results of the initial representation analysis for the down-sampled 1997 input profiles. It can be observed that for both the CAQE and CAQE ramps and for both the basic statistics and hybrid methods the trends align with those observed in the initial representation analysis in 2008.

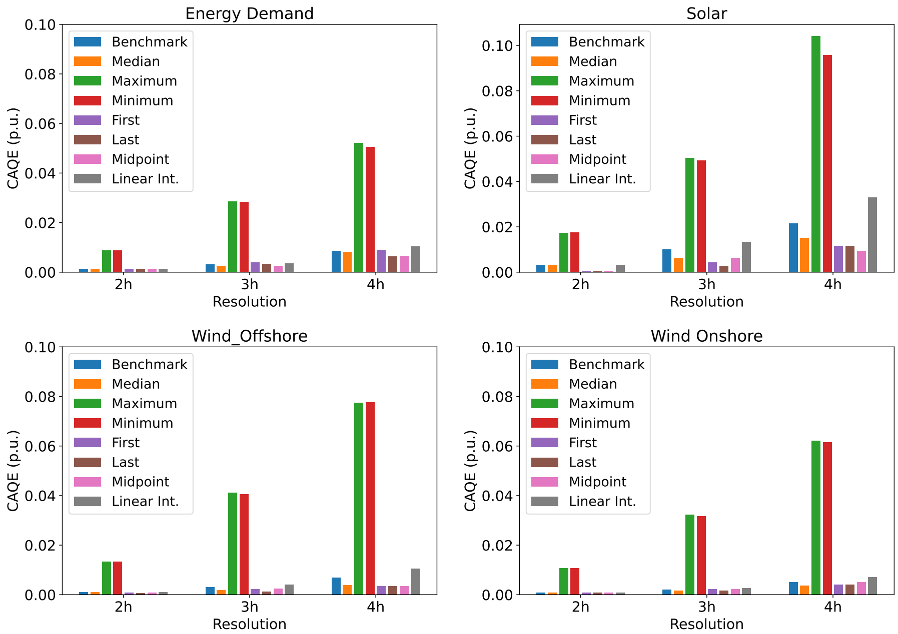


Figure L.53: Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

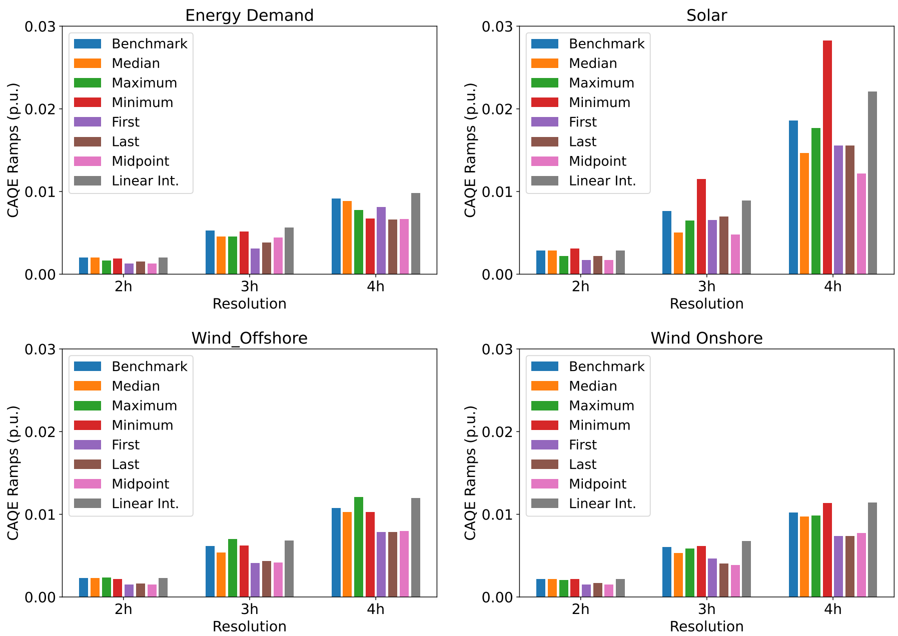


Figure L.54: Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with basic statistics methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

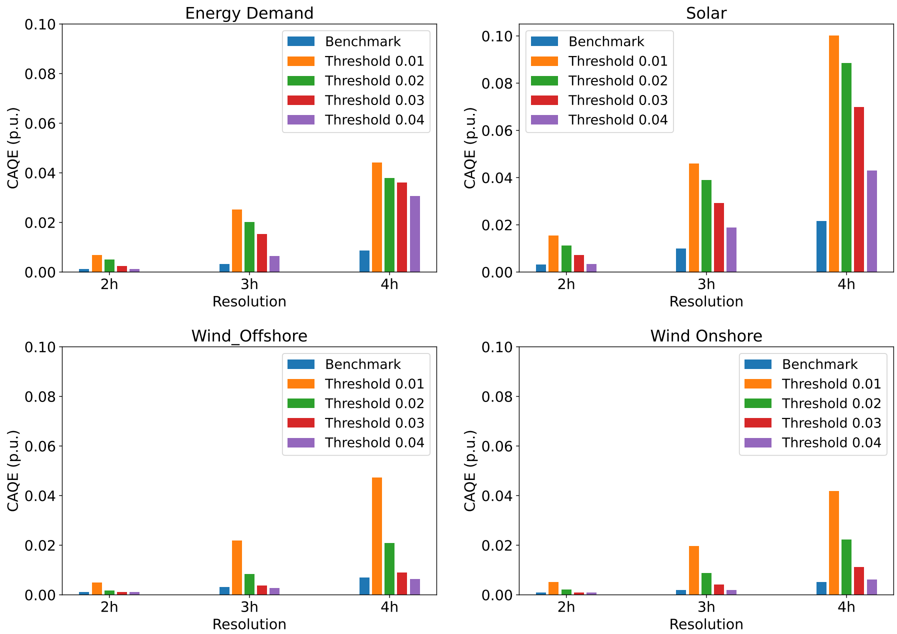


Figure L.55: Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of energy values for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid methods, to 2, 3 and 4-hourly resolution as indicated on the x-axis.

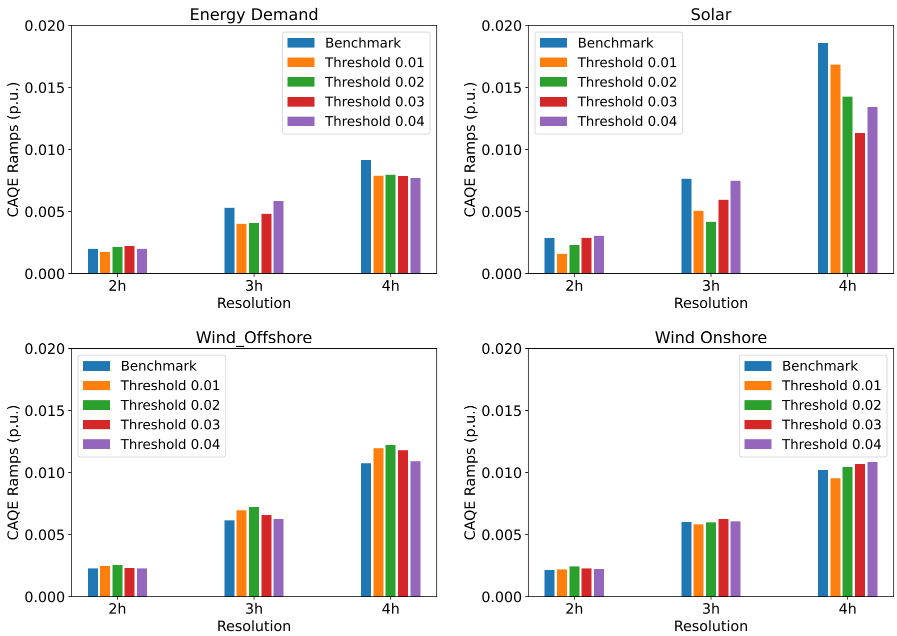


Figure L.56: Initial representation analysis results for 1997. Cumulative absolute quantile error of the distribution of ramps for the energy demand profile (top left), solar profile (top right), wind offshore (bottom left) and wind onshore profiles (bottom right). The profiles were down-sampled with hybrid method, to 2, 3 and 4-hourly resolution as indicated on the x-axis.