March 2025

Using storm classification for the improvement of mid-term forecasting of extreme storms along the Dutch coast

Prepared by: D.J. Najda

Hydraulic Engineering Section
Faculty of Civil Engineering and Geosciences
Delft University of Technology

The Hague, March 2025

Report type: MSc thesis report
Report name: Using storm classification for the improvement of mid-term forecasting of extreme storms along the Dutch coast
Cover image: Generated using AI-based image creation tools

Delft, university of technology (TUD)
Faculty of Civil Engineering
Department of Hydraulic Engineering

By	David. J. Najda	(TUD 4651081)
Graduation Committee	Dr. ir. J.A.Á. Antolínez	(TUD)
	Dr. ir. O. Morales-Nápoles	(TUD)
	ir. P.E. Kindermann	(TUD/HKV)
	ir. J.J. Caspers	(HKV)
Contact	David. J. Najda	

Preface

This thesis before you is part of my Master of Science in Civil Engineering at Delft University of Technology. It marks the end of my studies, a significant milestone in my academic journey. I began this research in September 2024, and it has been a rewarding and challenging experience that has brought me to the completion of this thesis in March 2025.

The completion of this thesis would not have been possible without the support of several people.

First and foremost, I would like to express my sincere gratitude to my graduation committee. To Dr. ir. J.A.Á. Antolínez, for his invaluable insight into the topic of weather pattern classification. His guidance in explaining the state-of-the-art research and particularly in helping me structure my approach while developing self-clustered weather patterns was incredibly helpful. Dr. ir. O. Morales-Nápoles, for his thoughtful contributions and assistance with the statistical aspects of the evaluation. I am also deeply grateful to Ir. P.E. Kindermann and Ir. J.J. Caspers, for their continuous support throughout this project. Our weekly meetings were always a great source of inspiration and fresh ideas, and their feedback consistently helped refine my approach and pushed my thinking forward. Ir. P.E. Kindermann also deserves special thanks for giving me the opportunity to work on this topic, setting the course for this research from our first conversations.

I also wish to express my gratitude to Ir. R.P. Verboeket for the warm welcome and informative tour during my visit to Watermanagementcentrum Nederland (WMCN). Furthermore, my thanks go to Ir. R. Neal from the Met Office for answering my questions on his research methods and for generously sharing data with me that supported my work. A special thanks goes to the entire team at HKV lijn in water where I carried out my research. Their resources, expertise, and welcoming environment were invaluable to me throughout this process.

Finally, I want to express my heartfelt thanks to my mother, Irina Najda, whose unwavering support and belief in me has been a constant throughout my life. Her gentle encouragement has not only helped me through my academic journey, but has been a pillar of strength in everything I do. I could not have done this without her.

I hope you find this thesis both informative and thought-provoking. Happy reading!

David Jonathan Najda
The Hague, March 2025

Abstract

Storm surges pose a critical threat to the Dutch North Sea coast, where low-lying areas are highly vulnerable to extreme water levels. While short-term storm surge forecasting is well established, extending predictions beyond 10 days remains a challenge due to the complexity of atmospheric dynamics and limitations in numerical models, which require extensive computational resources and lack flexibility for alternative forecasting techniques. This study explores the potential of weather pattern-based classification as a method for improving mid-term storm surge prediction. Phase I evaluates a set of predefined weather patterns (Neal, et al., 2018), originally developed by the Met Office for probabilistic forecasting in the UK, to assess their applicability for surge forecasting along the Dutch coast. The results indicate that while certain patterns show associations with high-surge events, they systematically underestimate surge magnitudes and, at longer lead times, the surge distributions associated with different patterns become increasingly similar, reducing their ability to distinguish between high- and low-surge conditions. Phase II explores an alternative self-clustered classification approach using k-means clustering with Principal Component Analysis (PCA) for dimensionality reduction. Several data selection methods, including surge thresholding, Maximum Dissimilarity Algorithm (MDA), and stratified sampling, are tested to optimise clustering. While the self-clustered patterns show slight improvements over Neal's predefined patterns, they still underestimate surge magnitudes and lack the accuracy needed for operational forecasting. A proof-of-concept evaluation using storm Pia (December 2023) and a representative SEAS5 storm reveals that the self-clustered weather patterns struggle to capture extreme surge events. Although these methods are not yet suitable for operational forecasting, this study suggest several possible refinements, such as sequential clustering and expanding the spatial domain, to be promising avenues for enhancing predictive skill.

Summary

Storm surges pose a critical threat to coastal regions, particularly along the Dutch North Sea coast, where low-lying areas are highly vulnerable to extreme water levels. While short-term storm surge forecasting (up to five days) is well established, mid-term forecasting (beyond 10 days) remains a challenge. This difficulty arises not only from the complexity of atmospheric dynamics and ocean-atmosphere interactions but also from limitations in numerical forecasting models. Current operational surge forecasts rely on high-resolution physical models that simulate surge behaviour based on meteorological conditions. While these models offer high reliability within a one-week timeframe, extending forecasts beyond 10 days is constrained by computational demands, as long-range simulations require extensive processing power, and model rigidity, as existing models are pre-calibrated and lack the flexibility needed to test alternative forecasting techniques.

At the same time, the growing need for mid-term storm surge forecasting is driven by maintenance planning for Dutch flood defences. With many storm surge barriers and dikes aging, Rijkswaterstaat aims to distribute repair operations more evenly throughout the year, rather than concentrating them in summer when storm risk is lower. Reliable mid-term predictions could help optimise these schedules while ensuring flood defences remain functional with minimal disruptions.

This study investigates the potential of weather pattern-based classification as a complementary tool for identifying high-risk periods earlier, providing an early indication of surge conditions at mid-range lead times to support preparedness and planning.

The research is divided into two phases. Phase I assesses the viability of using predefined weather patterns (Neal, et al., 2018), which were originally developed by the Met Office for classifying large-scale atmospheric circulation over the UK and surrounding region. This phase investigates whether these same patterns can be applied to storm surge forecasting along the Dutch coast by matching them to SEAS5 atmospheric fields. The results indicate that while certain patterns show some association with high-surge events, their overall predictive skill remains too weak for operational implementation. Additionally, at extended lead times, the surge distributions associated with different patterns become increasingly similar, reducing their ability to differentiate between high- and low-surge conditions.

Phase II explores an alternative self-clustered classification approach using k-means clustering in combination with Principal Component Analysis (PCA) as a dimensionality reduction technique to generate new weather patterns tailored to Dutch storm surges. Several data selection methods, including surge thresholding, Maximum Dissimilarity Algorithm (MDA), and stratified sampling, were tested to optimise pattern separation. An entropy analysis is conducted to further assess the effectiveness of these patterns. The results showed that while self-clustered patterns provided slightly better differentiation than Neal's predefined patterns, they retained considerable within-pattern variability, limiting predictive skill. At longer lead times, patterns became less distinct, reinforcing their decreasing predictive value over time.

A proof-of-concept evaluation is conducted using storm Pia (December 2023) and a representative SEAS5 storm to assess real-world applicability. The findings indicate that the self-clustered weather patterns systematically underestimate surge events across all lead times, reducing their reliability as a forecasting tool in their current form.

This study demonstrates that while the tested weather pattern-based classification methods are not yet suitable for operational mid-term surge forecasting, they highlight avenues for improvement. Further refinements, such as sequential clustering and expanding the spatial domain, may enhance predictive skill and improve classification accuracy.

Table of Contents

1
Introduction

1.1 Background

The Dutch coastal area is characterised by its low-lying topography, with significant portions of the land situated below sea level. This makes the Netherlands particularly vulnerable to coastal flooding. Ensuring the reliability of flood forecasting models is imperative to anticipate the arrival, timing, and intensity of these storms. While short-term forecasts (around one week ahead) facilitate timely operational decisions such as closing storm surge barriers, mid-term forecasting becomes crucial for maintenance and post-storm repair planning. Currently, the Water Management Centre Netherlands (WMCN) provides water level forecasting with a lead time of 10 days along the Dutch coast. However, there is an increasing demand for longer lead-time forecasts, similar to those already existing for the Dutch main rivers (Sumihar & Muis, 2024).

1.2 Dutch coastal flood defences

To protect the hinterland from the threat of inundation, the Netherlands has developed an extensive system of flood defences. These include dikes, dunes, and storm surge barriers, which are strategically designed and constructed to withstand extreme weather conditions and prevent seawater from breaching the coastline. Notable examples include:

- The Maeslant barrier (a movable flood barrier comprising two massive gates which protect densely populated areas, including the city of Rotterdam and its vital port).
- The Eastern Scheldt barrier (a partly movable barrier protecting the Eastern Scheldt estuary while allowing tidal flow into the estuary to maintain the marine ecosystem).
- The Afsluitdijk (a 32-km long dike and causeway system that separates the North Sea from the IJsselmeer, protecting a significant portion of the Dutch low-lying areas from flooding as well as facilitating land reclamation and agriculture) (Steenhuis, 2015).

Figure 1-1 shows a map of the Netherlands, highlighting these flood defence structures.

Figure 1-1: Coastal flood defence infrastructure in the Netherland. Image adapted from (Rijkswaterstaat, 2009).

1.3 Relevance of mid-term forecasting

Accurate storm surge forecasting is essential for coastal flood preparedness and operational decision-making in the Netherlands. The Water Management Centre Netherlands (WMCN) provides short-term storm surge forecasts (0–10 days) along the Dutch coast, enabling authorities to take necessary precautions such as:

- Closing storm surge barriers (e.g., Maeslantkering, Eastern Scheldt Barrier).
- Reinforcing flood defences where needed.
- Issuing warnings to water managers and emergency response teams.

These forecasts rely on physical models that simulate storm surge behaviour based on meteorological conditions, offering high reliability within a one-week timeframe. However, extending forecasts beyond 10 days remains a significant challenge due to several factors:

- The ***computational demands*** of high-resolution physical models, which require extensive processing power, making long-range forecasts time-intensive.
- The ***lack of flexibility*** in existing models, which are pre-calibrated, making them difficult to adjust when testing alternative forecasting techniques.
- The ***inherent uncertainties*** in mid-term atmospheric and oceanic conditions, leading to reduced forecast accuracy at longer lead times.

Despite these limitations, there is a growing demand for improved mid-term storm surge forecasting, particularly for maintenance and repair planning of flood defences. Many Dutch storm surge barriers and dikes are aging, requiring more frequent maintenance to remain operational. Traditionally, maintenance activities were concentrated in the summer months, when the risk of severe storms was lower. However, due to increasing maintenance demands, Rijkswaterstaat aims to spread repair operations more evenly throughout the year.

For this to be feasible:

- Accurate mid-term forecasts are needed to determine whether maintenance activities can proceed safely.
- If an extreme storm is expected, maintenance must be halted, postponed, or accelerated to ensure flood defences remain functional.
- Optimised resource allocation could reduce disruptions and ensure that repairs are completed efficiently before a storm event.

Given these challenges, WMCN have expressed growing interest in alternative forecasting approaches that can complement existing physical models. This study contributes to this effort by evaluating whether integrating large-scale weather pattern classification into storm surge forecasting can improve predictive forecasting for the Dutch coast.

1.4 Problem statement

In recent years, several extreme storm events have shown the limitations of mid-term storm surge forecasting. Notable examples include:

- ***Storm Eunice and Storm Franklin (2022)*** – Mid-term forecasts underestimated storm intensities, affecting flood preparedness and response efforts (Zijderveld, Verboeket, Bosma, & IJpelaar, 2022).

- ***St. Jude's Day Storm (2013)*** – The Met Office issued a five-day lead time warning for the UK, yet uncertainties in track and intensity made it difficult to anticipate the storm's full impact (Met Office, 2013).
- ***Storm Xaver (2013)*** – Initial forecasts predicted a "significant risk" of flooding 5–6 days ahead, but some coastal regions experienced more severe impacts than expected (Wadey, et al., 2015).

These cases illustrate a gap in mid-term storm surge forecasting, where predictions may lack the necessary lead time or precision for effective flood management.

A potential way to address this challenge is weather pattern-based storm classification, which has been explored in operational forecasting systems such as the Coastal Decider developed by the Flood Forecasting Centre (FFC) in the UK (Neal, et al., 2018). This approach classifies storm events based on predefined large-scale weather patterns in combination with high astronomical tides to estimate coastal flood risks. While this method has shown promise in the UK, its applicability to the Dutch coast remains unexplored.

1.5 Research objectives

The objective of this research project is to adapt and apply the Coastal Decider methodology developed by the Flood Forecasting Centre (FFC) in the United Kingdom to improve mid-term storm forecasting for the Dutch coast. By integrating large-scale weather pattern classification with nearshore hydraulic conditions, this study aims to enhance the accuracy of storm predictions beyond the traditional short-term forecasting horizon.

Beyond the application of these predefined weather patterns, this study also develops a self-clustered set of weather patterns tailored to the Dutch coast, using SEAS5 pressure fields. By comparing these newly derived patterns with the predefined set of Neal et al. (2018), the research assesses their respective performance in mid-term storm surge forecasting and explores their potential added value in improving predictive capabilities.

1.6 Research questions

The main research question of this study is:

To what extent can the integration of weather pattern-based storm classification contribute to improving mid-term forecasting accuracy for extreme storms impacting the Dutch coast?

This main question is further explored through the following sub-questions:

1. What are the key large-scale weather patterns associated with extreme storms impacting the Dutch coast and how do these identified weather patterns correlate with nearshore hydraulic conditions?
2. How well do the weather patterns from Neal et al. (2018) perform in predicting storm surges along the Dutch coast?
3. What is the added value of self-clustered weather patterns compared to Neal's predefined weather patterns in mid-term storm surge forecasting?
4. What are the potential practical implications and operational utility of integrating weather pattern-based storm classification to mid-term storm forecasting for coastal flood preparedness and response strategies in the Netherlands?

1.7 Research approach

To address the posed research questions, this study is structured into two distinct phases, each focusing on different aspects of weather pattern-based storm surge forecasting:

Phase I: Evaluating Neal's predefined weather patterns
The first phase assesses the applicability of Neal's predefined weather patterns to storm surge forecasting for the Dutch coast. SEAS5 pressure fields are matched to these patterns, and their relationship with extreme storm surges is analysed. This phase evaluates the effectiveness of Neal's patterns in capturing key atmospheric conditions associated with high surge events and determines their predictive value for mid-term storm forecasting.

Phase II: Developing and assessing self-clustered weather patterns
In the second phase, a new set of weather patterns is derived specifically for the Dutch coast using self-clustering techniques on SEAS5 pressure fields. This phase explores whether locally optimised weather patterns provide improved predictive skill over Neal's predefined set. The clustering process, classification performance, and the relationship between these patterns and coastal storm surges are examined to assess their added value in storm forecasting.

1.8 Significance of the study

This study contributes to both scientific knowledge and operational decision-making by:

- Evaluating whether pattern-based classification can enhance mid-term storm forecasting.
- Providing WMCN with insights into potential alternative forecasting methods.
- Exploring how data-driven approaches could complement existing physical models for storm surge prediction.

The findings may help inform future developments in storm surge forecasting, ultimately improving flood preparedness and risk management in the Netherlands.

1.9 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2: Literature review provides an overview of the theoretical background relevant to this study. It includes a discussion on storm formation, storm surge dynamics, coastal flood risks, and forecasting methodologies, with a particular focus on weather pattern-based forecasting. Additionally, this chapter introduces the case study region: the Dutch North Sea coast.

Chapter 3: Relevant datasets describes the meteorological and hydrodynamic datasets used in this research. It details data sources such as SEAS5, DCSM5, EMULATE and ERA5, explaining their relevance and role in storm surge forecasting.

Chapter 4: Exploratory analysis SEAS5 examines the SEAS5 dataset, providing an initial analysis of its structure, key variables, and its potential for mid-term forecasting. Considerations related to initial conditions and surge behaviour are also explored.

Chapter 5: Research method phase I outlines the methodology for applying Neal's predefined weather patterns to SEAS5 atmospheric fields. It details data preprocessing, matching algorithms, and the l inking of surge events to specific weather patterns.

Chapter 6: Evaluation of Neal weather patterns presents the results of Phase I, analysing the suitability of Neal's predefined weather patterns for storm surge forecasting along the Dutch coast. It includes an assessment of pattern-matching accuracy, surge distribution, and predictive performance.

Chapter 7: Research method phase II describes the approach for generating self-clustered weather patterns using k-means clustering. This includes data selection, clustering methods, and dimensionality reduction techniques.

Chapter 8: Data selection explores different methods for selecting relevant surge cases for clustering. The methods include surge thresholding, the Maximum Dissimilarity Algorithm, and stratified sampling. Their advantages, limitations, and impact on weather pattern classification are discussed.

Chapter 9: Evaluation of generated weather patterns presents the results of Phase II, evaluating the newly generated weather patterns in their effectiveness in distinguishing storm surge variability and their potential improvements over existing forecasting approaches.

Chapter 10: Evaluation of forecasting capability: a proof of concept explores the predictive performance of the weather pattern classification methods in a real-world forecasting scenario. Using storm Pia and a representative SEAS5 storm, this chapter assesses the practical viability of pattern-based surge forecasting and identifies key strengths and limitations of the approach.

Chapter 11: Discussion interprets the findings from both research phases, providing an evaluation of the methodology and its implications. It addresses theoretical expectations versus actual performance, highlights methodological limitations, and discusses potential interdisciplinary applications of weather pattern-based forecasting.

Chapter 12: Conclusion and recommendations summarises the main findings of the research and summarises the key finding of the study and discusses their broader significance. This chapter concludes with recommendations for future research and potential practical considerations.

Appendices include additional details on supporting methodologies and supplementary analyses.

2
Literature review

2.1 Storm formation and development

Storm formation along the Dutch coast in the North Sea is influenced by a convergence of atmospheric and oceanic factors. Cold polar air collides with warm, moist air from lower latitudes, creating a sharp temperature gradient. This contrast leads to the development of an instability along the polar front, initiating depression formation (Kraker, 2002). Typically, within 24 hours this instability evolves into a fully-developed cyclone, characterised by a warm front extending north-eastward and a cold front trailing south-eastward. The area between these fronts, known as the "warm sector," experiences overrunning of warm air and the surging of cold air. As the cyclone intensifies, wind speeds increase and precipitation becomes widespread. Eventually, the storm matures, with the cold front overtaking the warm front, resulting in the formation of an occlusion front. The storm gradually dissipates, losing energy as it moves over land or cooler waters (Aguado & Burt, 2001; Northern Vermont University, n.d.). Figure 2-1 schematically depicts the formation and development of such a polar front depression.

The instability along the polar front is further amplified by the presence of the North Atlantic storm track, which often directs powerful low-pressure systems towards the region. As these low-pressure systems move over the relatively warm waters of the North Sea, they intensify, leading to the development of strong winds, heavy rainfall, and storm surges along the Dutch coast (Bell, Gray, & Jones, 2017).

In addition to these large-scale drivers, atmospheric blocking can play a significant role in modulating storm development and storm surge dynamics. Blocking high-pressure systems over Scandinavia or Central Europe can obstruct the usual west-to-east progression of weather systems, causing cyclones to stall or redirect over the North Sea. This stagnation can lead to sustained onshore winds and prolonged high sea levels along the Dutch coast, increasing the risk of extreme storm surges (Kautz, et al., 2022). Such blocking patterns can effectively alter storm tracks, either intensifying or prolonging adverse weather conditions, thereby shaping the severity of storm surge events in the region. A notable example of such an event was Storm Xaver in December 2013, where a strong anticyclone over southwestern Europe created a persistent and intensified pressure gradient across the North Sea. This blocking setup led to sustained north-north-westerly winds, driving record-breaking storm surges along the German Bight and impacting the Dutch coastline as well (Kautz, et al., 2022).

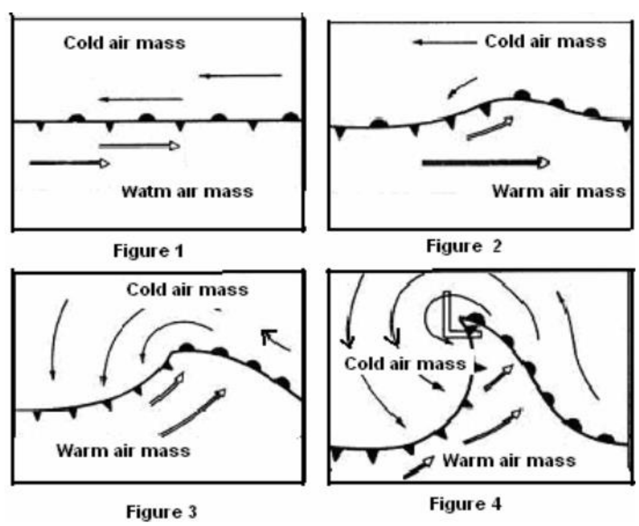


Figure 2-1: The formation and development of a polar front depression. Panel 1: Warm, moist air masses from the south meets cooler air masses from the north along the frontal boundary. Panel 2: A disturbance develops on the frontal boundary between the two air masses. Panel 3: The warm front extends north-eastward and the cold front trails south-eastward. Panel 4: The depression is fully formed. Figure taken from (Chowdhury, 2017).

Winter along the Dutch coast brings about a higher likelihood of extreme weather conditions, with winter storms typically being associated with stronger baroclinic zones, characterised by sharp temperature contrasts. These contrasts are more pronounced in winter due to the larger difference in temperature between the poles and the equator (Schwierz, et al., 2010).

2.2 Storm surge dynamics

Storm surge is defined as an abnormal rise in sea level along a coast and is primarily caused by the intense winds and low atmospheric pressure associated with storms. Changes in atmospheric pressure produce immediate effects on the forces acting vertically on the sea surface, impacting sea level at all depths, while wind stress generates forces parallel to the sea surface (Pugh, 1987). Strong onshore winds push water towards the shore, resulting in sea levels rising above predicted tide levels (Holthuijsen, 2007). The effects of winds and air pressures are often intertwined in storms, making it challenging to separate their individual impacts. In extratropical storms like those occurring in the North Sea, both pressure and wind effects may be equally important, unlike tropical storms where wind stress typically dominates (Pugh, 1987).

The space and time scales of storm surges closely mirror those of the generating storm, typically spanning a few hundred kilometres and lasting one to two days (Holthuijsen, 2007). However, responses to weather forces may vary, particularly in semi-enclosed seas like the North Sea, where surges can persist for more than one tidal cycle (Pugh, 1987). The extensive spatial scales and longer durations of extratropical storms mean that the effects of the Earth's rotation, represented by Coriolis forces, play a significant role in determining the seas' dynamical response, by diverting wind driven currents towards or away from the coast (Byrne, Horsburgh, Zachry, & Cipollini, 2017). Additionally, the natural resonant periods of seas and basins themselves influence the behaviour of storm surges (Pugh, 1987).

Factors influencing storm surge severity

Based on a comprehensive analysis of past historical flood events along the Dutch coast by de Kraker (2010), a trend in meteorological conditions contributing to the severity of storm surges and flooding events in the Netherlands has been observed. These conditions include:

I. ***Wind direction*** – Storm surges associated with westerly and north-westerly gales pose significant risks to the Dutch coast. For example, the storm surge of 1 February 1953 was associated mostly with a north-westerly wind direction blowing from the North Sea.
II. ***Moon phase*** - Storm surges coinciding with the Full Moon or New Moon phases tend to be more severe due to the gravitational influence on tides. Examples of this include the floods of 1953, 1906 and 1808, which all occurred during periods of high spring tides.
III. ***Duration*** - Longer-duration storm surges, spanning several tidal cycles, can exacerbate flooding risks. For instance, the flood event of 1953 lasted for approximately two days, during which three consecutive high spring tides caused unparalleled inundation levels.
IV. ***Intensity*** - Storm surges with high wind speeds can lead to more extensive flooding and damage. During the storm surge of 1953, high wind speeds of force 10–11 on the Beaufort scale were observed, caused widespread devastation across the Dutch coastal region.

Storm surge and tidal interactions

Another important aspect to consider is the interaction between storm surge and tides, particularly when it comes to predicting coastal water levels. The timing of the storm surge relative to the astronomical tide has a significant impact on the actual observed water levels along the coast. Previous studies, such as Horsburgh & Wilson (2007) and Geerse (2020) have investigated these interactions and found that storm surges often do not coincide with the highest astronomical tide due to several reasons, namely:

I. ***Phase shift*** - The interaction between tide and surge involves a mutual phase alteration. Phase shift refers to the change in the timing of the tidal cycle due to the presence of a storm surge. Essentially, a storm surge can either advance (move forward) or retard (move backward) the timing of the peak of the tide. A positive storm surge (an increase in sea level) can increase the speed at which the tidal wave travels. This is because the surge adds to the water depth, and waves travel faster in deeper water. Conversely, a negative surge (a decrease in sea level) can slow down the tidal wave because it reduces the water depth.

II. ***Tidal modulation*** - The generation of a storm surge is modulated by the tidal state. During low water, the water depth is shallower, which makes it easier for wind stress to raise the sea surface. This effect is less pronounced at high water when the water depth is greater. As a result, the surge is often larger at low water than at high water. This phenomenon is known as tidally modulated surge production.

III. ***Traveling surge components*** - Surges can travel along the coast, and their timing relative to the tidal cycle can vary. When a surge is generated at a different location, it may arrive at a particular site at a different time relative to the tidal cycle than if it had been generated locally. This traveling nature of surge components means that the peak of the surge may not coincide with the peak of the astronomical tide when it reaches a specific location.

2.3 Coastal flood risk

The Netherlands faces a significant coastal flood risk due to its low-lying geography, with large areas situated below sea level. Coastal storm surges, driven by extreme weather events, pose a continuous threat to flood defences, infrastructure, and human safety. To manage these risks, the Dutch flood protection system relies on a combination of engineering measures, operational forecasting, and crisis management strategies (enw | expertisenetwerk waterveiligheid, 2017).

Floor risk governance and institutional responsibilities

In the Netherlands, the responsibility for decision-making regarding land protection against flooding falls primarily to Rijkswaterstaat and local water boards. These entities are tasked with the maintenance and operation of flood defence systems, as well as coordinating responses to potential flood threats. The operational teams within these organisations requiring timely and accurate information to make informed decisions about actions such as closing storm surge barriers. This is vital for ensuring the safety and security of the Dutch hinterland, where timely intervention can significantly mitigate the impact of coastal flood events.

The strategy for coordinating responses to potential flood threats involves a collaborative effort between various parties, including (WMCN, 2023; enw | expertisenetwerk waterveiligheid, 2017):

- **Rijkswaterstaat (RWS) -** The primary agency responsible for the maintenance and operation of national flood defences, including storm surge barriers, dikes, and dunes. RWS also plays a role in storm surge forecasting and crisis coordination.
- **Watermanagementcentrum Nederland (WMCN) -** The national center for water management, responsible for storm surge forecasting, issuing flood warnings, and advising decision-makers during flood events. WMCN consists of multiple operational teams, including the Waterkamer, which monitors national water systems and issues warnings, and the Landelijke Coördinatiecommissie Overstromingsdreiging (LCO), which is activated when a (potential) flooding threat arises. In the lead-up to this, the Waterkamer and crisis advisory groups are already actively engaged in monitoring and decision-making.
- **Koninklijk Nederlands Meteorologisch Instituut (KNMI) -** Provides meteorological forecasts and collaborates with WMCN to predict storm surges and high water events.

- **Water Boards (Waterschappen) -** Manage regional water infrastructure, oversee dike surveillance, and implement localised emergency measures. They are collectively represented by the Unie van Waterschappen (UvW).
- **Departementaal Coördinatiecentrum Crisisbeheersing (DCC-IenW) -** The crisis management unit within the Ministry of Infrastructure and Water Management (IenW). It coordinates national-level decision-making during major flood threats and serves as a link between the water crisis management structure and broader national crisis management organisations.
- **National Crisis Center (NCC) -** Activated in extreme scenarios, leading national-level crisis management and public safety coordination.
- **Veiligheidsregio's (Safety Regions) -** Regional emergency response units responsible for public safety, evacuations, and coordination with local municipalities during flood events.

Flood warning system and colour coding

Forecasts for storm surges are issued jointly by RWS and KNMI. When high water levels are expected, WMCN takes over coordination and advises local authorities on actions to be taken. The accuracy of storm surge forecasts is imperative for optimal operation of storm surge barriers and guidance for local authorities, with preparation benefiting from forecasts up to 10 days ahead (de Vries, 2013). The warning system employs color-coded alerts to communicate the threat posed by high water. This system, outlined in WMCN (2023) ensures that all involved authorities, from water managers to crisis coordinators, operate with a clear understanding of the severity of an impending flood event. Table 2-1 summarises these alert levels.

Table 2-1: Flood risk warning levels in the Netherlands.

Warning Level	Description	Actions
Green	Normal conditions	- No imminent flood risk. - Routine monitoring by water authorities. - No special measures required.
Yellow	Increased awareness	- Potential flood threat detected, requiring heightened readiness. - Water levels are expected to rise but remain manageable. - Water authorities take standard pre-emptive measures. - Can occur multiple times a year.
Orange	High alert	- Serious flood risk. - Authorities take further pre-emptive measure. - Occurs on average once every 5 years.
Red	Extreme flood risk	- Severe flooding is imminent or occurring. - Full-scale crisis response measures are implemented. - Occurs on average once every 20-100 years.

These warnings are based on the probability of reaching certain predefined levels of storm surge, as outlined in Table 2-2 by de Vries (2013). When the probability of reaching at least the ‘Warning’ level within 8 days in any coastal region exceeds 25%, KNMI contacts WMCN, triggering an escalation ladder for coordination of actions to be taken to deal with the possible consequences of high water.

Table 2-2: Significant levels for storm surge forecasts (de Vries, 2013)

Level	Action	Exceedance $[year^{-1}]$
Information	KNMI informs WMCN	10
Pre-warning	WMCN issues limited warnings	5
Warning	WMCN office opened, issues warnings	2
Regional Alarm	WMCN advises local authorities, LCO active	0.2
National Alarm	DG-RWS leading	$5 * 10^{-2} - 10^{-2}$
Critical	NCC leading	$5 * 10^{-3} - 10^{-3}$

2.4 Forecasting methods

Storm surge forecasting is important for predicting and preparing for extreme storm events and safeguarding coastal communities from potentially devastating inundation. To achieve accurate predictions, two main approaches are commonly used: numerical modelling and statistical modelling. This section explores these approaches and various methodologies within each category, drawing from past studies and literature research.

Numerical modelling vs. statistical modelling

Numerical modelling involves solving physical equations to simulate storm surge dynamics, while statistical modelling establishes relationships between atmospheric variables and storm surge outcomes using historical data.

- ***Numerical weather prediction (NWP)* -** Numerical weather prediction (NWP) involves the use of computer models to simulate atmospheric processes and predict future weather conditions (von Storch, 2014). These models utilise equations of motion and mass conservation to forecast atmospheric variables such as wind speed, air pressure, and precipitation. NWP provides input data for storm surge models by predicting the meteorological conditions that drive storm surge events (Li & Nie, 2017).
- ***Statistical models* -** Statistical models utilise past data to identify patterns and relationships between meteorological variables and storm surge outcomes. These models, while simple and fast, may suffer from over- or under-specification, limiting their robustness and ability to account for changing geophysical conditions, such as global climate change or changes to the near shore topography due to for instance engineering works. Such limitations are inherent to the method, given its reliance on statistics from past events (von Storch, 2014).

The main advantage of numerical models is that they provide detailed and high-resolution representations of physical processes. Additionally, they can predict storm surges at locations without direct observations, a significant advantage over purely statistical methods (Woth, 2005). However, they are computationally expensive and require extensive data for boundary conditions and accurate local modelling (Costa, Idier, Rohmer, Menendez, & Camus , 2020). The main advantages of statistical models are their simplicity and low computational cost, making it feasible to run multiple simulations for long-term predictions (Costa, Idier, Rohmer, Menendez, & Camus , 2020; Woth, 2005). However, they often require extensive historical data and may struggle with short-term predictions and extreme events, especially in regions with complex geography like enclosed seas and bays (Costa, Idier, Rohmer, Menendez, & Camus , 2020).

Examples of numerical modelling methods

I. Hydrodynamic models

Hydrodynamic models simulate currents and water levels in coastal regions by solving equations of motion and mass conservation (von Storch, 2014). These models, such as the Dutch continental shelf model (DCSM), use numerical methods to discretise equations and incorporate boundary conditions from meteorological forecasts (Verlaan, Zijderveld, de Vries, & Kroos, 2005). By considering factors such as wind stress, atmospheric pressure, and Coriolis force, hydrodynamic models predict storm surge dynamics (Li & Nie, 2017).

II. Ensemble forecasting

Ensemble forecasting involves running multiple simulations with variations in initial conditions, boundary conditions and model physics to account for the range of uncertainty in model inputs and formulation (Flowerdew, Horsburgh, & Mylne, 2009). By generating a range of possible outcomes, ensemble forecasting provides probabilistic forecasts. This approach contrasts with traditional single-model forecasting methods by explicitly considering the variability and uncertainty inherent in the atmospheric conditions driving storm surge events (Flowerdew, Horsburgh, & Mylne, 2009). Ensemble forecasts offer several advantages, as highlighted by Kohno et al. (2018). They provide various storm surge values such as mean, maximum, and minimum, allowing for a comprehensive assessment of potential outcomes. Additionally, ensemble forecasts offer spread information, aiding in the interpretation of reliability and the range of predicted values. However, achieving this higher reliability requires a larger number of forecast members, which in turn demands more computational resources.

III. Data assimilation techniques

Data assimilation techniques integrate real-time observational data, such as wind field datasets (Byrne, Horsburgh, Zachry, & Cipollini, 2017), satellite altimetry data (Madsen, Høyer, Fu, & Donlon, 2015), tide gauge measurements (Etala, Saraceno, & Echevarría, 2015) and atmospheric pressure observations (Toyoda, Fukui, Miyashita, Shimura, & Mori, 2021) into numerical models to improve forecast accuracy. However, as emphasised by Verlaan, Zijderveld, de Vries, & Kroos (2005), it is crucial to validate the data before assimilation. Incorporating erroneous observations can have a disproportionately detrimental impact on the forecast, outweighing the benefits of correct observations. If implemented correctly, these techniques help correct model biases and enhance the reliability of storm surge predictions, particularly during rapidly evolving storm events.

Examples of statistical modelling methods

I. Statistical downscaling

Statistical downscaling is a technique used to derive high-resolution climate or weather information from coarse-resolution global climate model outputs. It involves establishing statistical relationships between large-scale atmospheric variables (predictors) and local-scale surface variables (predictands) using observational data (Feddersen & Andersen, 2005). The main steps in statistical downscaling are:

i. Identifying relevant large-scale predictors from global climate model outputs that influence local climate variables of interest.
ii. Developing statistical models that relate the predictors to observed local climate variables using techniques like regression, weather typing, or neural networks.
iii. Applying statistical models to global climate model outputs to obtain downscaled local climate projections.

There are several approaches used in statistical downscale, including:

i. ***Regression-based approaches*** - These include methods like artificial neural networks (ANNs), which have been used for short-term local predictions and studying extreme events (Costa, Idier, Rohmer, Menendez, & Camus , 2020; Schoof, 2013).
ii. ***Analog methods*** - These methods search historical records for patterns matching current conditions to predict local climate, requiring long historical series (Schoof, 2013).

iii. ***Weather generators*** - These stochastic models generate plausible daily weather data sequences for impact modelling, useful when observational data is insufficient (Schoof, 2013).

iv. ***Weather-type approaches*** - These cluster atmospheric data into weather types for climate variability studies, often used for monthly and annual projections (Costa, Idier, Rohmer, Menendez, & Camus , 2020; Rozas Rojas, 2017; Camus, et al., 2014; Schoof, 2013).

II. Deep learning / Neural Networks based forecasting

These methods utilise artificial intelligence techniques, such as deep learning algorithms or neural networks, to learn complex relationships between input meteorological data and storm surge outcomes, enabling data-driven forecasting approaches. Several studies have utilised different types of neural network architectures for storm surge forecasting, including Convolutional Neural Networks (CNN) (Xie, Xu, Zhang, & Dong, 2023), Long Short-Term Memory (LSTM) (Bai & Xu, 2022), Artificial Neural Networks (ANN) (Ramos-Valle, Curchitser, Bruyère, & McOwen, 2021) and Neural Networks (NN) (Lee, 2006). These methods offer the flexibility to capture nonlinear relationships in the data and can improve forecast accuracy, especially when traditional methods may struggle to account for complex interactions between meteorological and oceanographic variables. Additionally, deep learning models often exhibit faster computational speeds compared to their numerical counterparts, enhancing efficiency in storm surge prediction tasks.

Summary forecasting methodologies

Table 2-3 provides an overview of the presented storm surge forecasting methodologies and their respective inputs and outputs.

Table 2-3: Overview of storm surge forecasting methodologies

Methodology	Input	Output
Numerical Weather Prediction (NWP)	Atmospheric variables (e.g., wind speed, air pressure, temperature, humidity, precipitation), sea surface temperature, wind stress, atmospheric pressure gradients	Meteorological conditions driving storm surge events, including wind fields, atmospheric pressure patterns, and precipitation forecasts
Hydrodynamic Models	Meteorological conditions (e.g., wind stress, atmospheric pressure), bathymetry/topography, tidal forcing, land-sea boundary conditions	Currents, water levels, storm surge, inundation map, wave heights, coastal erosion rates
Ensemble Forecasting	Variations in initial conditions (e.g., atmospheric pressure, wind speed, temperature), boundary conditions (e.g., sea surface temperature, land-sea friction), model physics	Probabilistic forecasts of storm surge magnitude and extent, ensemble mean, maximum, and minimum values, spread information indicating forecast uncertainty
Data Assimilation Techniques	Real-time observational data (e.g., wind field datasets, satellite altimetry data, tide gauge measurements, atmospheric pressure observations), historical storm surge and meteorological data	Improved forecast accuracy through assimilation of observational data into numerical models, corrected model biases, enhanced reliability of storm surge predictions
Statistical Models	Historical storm surge data, historical meteorological data (e.g., wind speed, atmospheric pressure, storm track)	Probability of storm surge occurrence based on past events and meteorological conditions
Deep Learning / Neural Networks	Meteorological data (e.g., wind speed, atmospheric pressure, temperature), historical storm surge data, geographic features (e.g., bathymetry, shoreline characteristics)	Storm surge predictions based on learned relationships between meteorological variables and storm surge outcomes

2.5 Weather-pattern based forecasting

One way to enhance forecasts is through the use of pre-defined weather patterns, which efficiently assign ensemble members to the most suitable type, maximising information while minimising data quantity (Neal, Fereday, Crocker, & Comer, 2016). This section summarises the methodology employed in defining and generating weather patterns, as adapted from the work of (Neal, et al., 2018; Neal, Fereday, Crocker, & Comer, 2016). The approach uses cluster analysis techniques to delineate representative weather patterns over the UK and surrounding European area, making them suitable for application to the Dutch coastal context.

I. Data preprocessing

For the weather pattern generation, the EMSLP dataset was used, which provides a comprehensive record of daily mean sea level pressure (MSLP) fields spanning from 1850 to 2003 over the European and North Atlantic region (Allan, 2007). It should be noted that some biases exist in this dataset, particularly in regions with sparse data coverage or during summer months when local climatic conditions are harder to resolve. The dataset also tends to smooth out extreme values due to its gridding and infilling procedures (Ansell, et al., 2006). Before clustering, the smoothed climatology was removed from each field, thereby reducing the influence of the seasonal cycle on the pattern classification. Smoothed climatology refers to the long-term average or trend in the data, which can obscure short-term variations.

II. Cluster analysis choices

The cluster method involved several decisions to be made, namely:

- ***Single set vs. distinct sets*** - The choice was made to produce a single set of patterns for the entire year, as distinct sets for each season would have resulted in excessive pattern numbers and hindered inter-season comparisons.
- ***Number of patterns*** - A final set of 30 patterns was chosen, striking a balance between pattern richness versus similarity.
- ***Spatial domain choice*** - Through iterative evaluation and objective metrics, the domain (30°W–20°E; 35°–70°N) was identified as the most suitable, optimising temperature and precipitation reconstruction. This domain was primarily chosen for the UK area but also features the surrounding European area, making it suitable for use along the Dutch coast as well.

III. Cluster analysis

To analyse the pre-processed Mean Sea Level Pressure (MSLP), a simulated annealing variant of k-means clustering was employed. K-means clustering is a method used to partition a dataset into a predetermined number of clusters (denoted as k) based on similarity criteria. The process involves the following steps (Sharma, 2024):

i. ***Initialisation*** - Randomly select k data points from the dataset as initial cluster centroids.
ii. ***Assignment*** - Assign each data point in the dataset to the nearest cluster centroid based on a chosen distance metric, such as Euclidean distance.
iii. ***Update*** - Recalculate the centroids of the clusters based on the mean of all data points assigned to each cluster.
iv. ***Iteration*** - Repeat the assignment and update steps until convergence, i.e., until the centroids no longer change significantly, or a predefined number of iterations is reached.

Simulated annealing introduces a probabilistic approach to updating cluster assignments and centroids, allowing for robust exploration of the solution space. This method helps avoid local optima by occasionally accepting worse solutions based on a probability distribution, which can lead to improved clustering results compared to traditional k-means algorithms (Bandyopadhyay, Maulik, & Pakhira, 2001). For a more detailed explanation of the theory behind k-means cluster, refer Appendix B: Theory explained (Section B.2).

IV. Resultant weather patterns

The outcome of the clustering process yielded a set of 30 weather patterns, designed primarily for medium-range forecasting up to 15 days (see Figure 2-2). These patterns capture variations within broad-scale circulation types. The patterns are ordered according to their annual historic occurrence. Lower numbered patterns are more prevalent during the summer months, characterised by weaker MSLP anomalies. Conversely, higher numbered patterns are dominant in winter, exhibiting stronger MSLP anomalies.

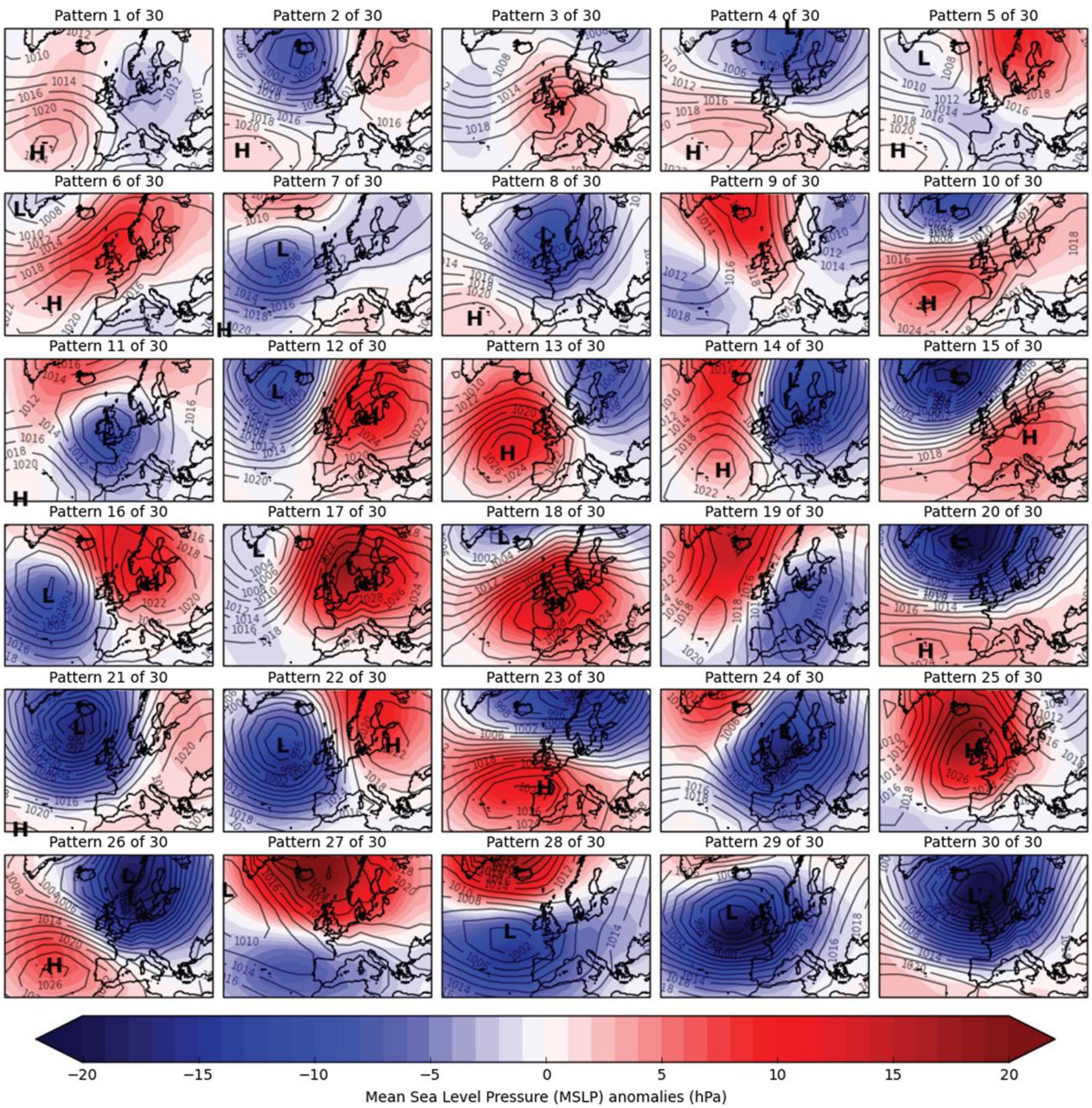


Figure 2-2: Weather patterns definitions. Figure taken from (Neal, Fereday, Crocker, & Comer, 2016).

2.6 Study area

2.6.1 North Sea

The North Sea, bordered by the United Kingdom, Norway, Denmark, Germany, Belgium, and the Netherlands, is a shallow shelf sea adjacent to the North Atlantic (see Figure *2-3*). What sets the North Sea apart, particularly in relation to flood risk, is its shallow depths and relatively narrow continental shelf, which amplify the impacts of extreme weather events (Sünderman & Pohlmann, 2011). The North Sea's proximity to densely populated coastal areas exacerbates the impact of flooding and coastal inundation, rendering coastal communities vulnerable to storm surges and their devastating consequences for lives, property, and critical infrastructure.

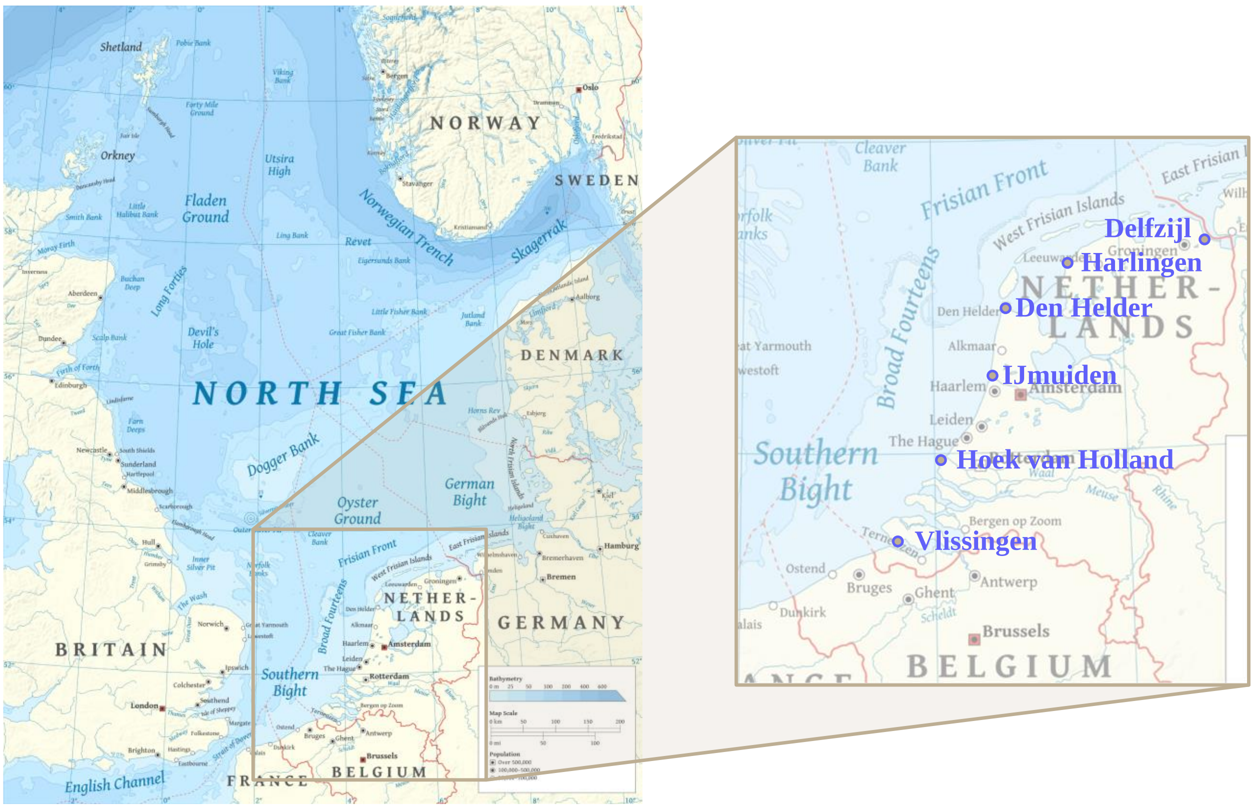


Figure 2-3: Map of the North Sea (left) and zoom-in of the Dutch coast (right). Image adapted from (Halava, 2015).

Tides

The dynamics of the North Sea are significantly influenced by astronomical tides resulting from the gravitational forces of the moon and sun acting on the Atlantic Ocean. Semidiurnal tides, particularly the M2 and S2 components, predominate at the latitudes concerned and are further amplified in the North Sea by resonance with the configuration of the coasts and depth of the seabed (Ozer & Legrand, 2015). These tidal oscillations contribute to the overall circulation and water mass transports within the North Sea.

Atmospheric dynamics

The atmospheric dynamics are important in shaping the circulation patterns and water mass development in the North Sea. Prevailing westerly winds on the northwest European shelf, associated with a meandering upper troposphere jet stream, contribute to the reinforcement of the cyclonic residual circulation induced by the tide (Ozer & Legrand, 2015). These winds, along with cyclonic activity embedded within the belt of westerly winds, are stronger in winter than in summer, further influencing the dynamics of the North Sea.

Storm surges

Storm surges constitute the most serious hazard in the North Sea region (Ozer & Legrand, 2015; Sünderman & Pohlmann, 2011). The North Sea's relatively shallow depths and its broad connection to the Atlantic Ocean allow for the free exchange of momentum, energy, and matter between the two seas, contributing to the development and propagation of storm surges (Sünderman & Pohlmann, 2011). Storm surges, once generated, travel in the same manner as tides, posing significant risks to coastal communities and infrastructure along the North Sea coastlines.

2.6.2 Dutch coast

The Dutch coast spans approximately 450 kilometres along the North Sea and features a diverse range of geographical features, including sandy beaches, dunes, tidal flats, and reclaimed land. Based on literature, this coastal stretch may be divided into three main areas: the Delta area, the coast of Holland and the Wadden area (van der Spek, et al., 2022; Stolk & Dillingh, 1989).

Delta area (Southwest)

The Delta Area is notable for its peninsulas separated by estuaries and former tidal basins, which have been shaped significantly by human interventions of the past aimed at flood protection and land reclamation (Stolk & Dillingh, 1989). The shoreface of the Delta Area consists of extensive ebb-tidal deltas with low-gradient platforms, dissected by tidal channels, and topped with intertidal to supratidal sand bars (van der Spek et al., 2022). Significant constructions in this area, include the storm-surge barrier in the Eastern Scheldt and the dams closing off the Veerse Gat, Grevelingen, and Haringvliet. These structures have transformed the region, reducing the influence of tidal processes and increasing the prominence of human-managed coastal features (Stolk & Dillingh, 1989).

Coast of Holland (West)

Stretching from Hoek van Holland to Den Helder, the Coast of Holland is characterised by a nearly continuous row of dunes with minimal interruptions from islands or tidal inlets. This region is primarily influenced by wave action rather than tidal forces due to the relatively steep gradient of the shoreface, which supports shore-parallel breaker bars that are heavily influenced by wave energy (van der Spek, et al., 2022; Stolk & Dillingh, 1989). Key infrastructural elements include the sea dikes at Hondsbossche Zeewering and Helderse Zeewering, and the ports of Scheveningen and IJmuiden, which are notable interruptions in the otherwise continuous dune system forces (Stolk & Dillingh, 1989).

Wadden Area (North)

The Wadden Area, extending from the north of the Netherlands to the German border, is characterised by its barrier islands and extensive tidal flats. This region is heavily influenced by tidal processes, with larger tidal ranges and numerous tidal inlets (Stolk & Dillingh, 1989). A major transformation in this region was the construction of the Afsluitdijk, a 30-kilometer-long dam completed in 1932, which closed off the Zuiderzee from the North Sea, turning it into the freshwater IJsselmeer. This project significantly altered the hydrodynamics of the Wadden area, reducing the tidal influence on the former Zuiderzee while protecting the inland regions from flooding (Stolk & Dillingh, 1989).

2.6.3 Considered locations

Six locations have been selected for this study: Delfzijl, Harlingen, Den Helder, IJmuiden, Hoek van Holland, and Vlissingen (see Figure *2-3*). These stations represent the six main tidal stations in the Netherlands, which are evenly distributed along the Dutch coast and have a long history of measurements (Stolte, et al., 2023). Most of these stations have been continuously measuring the water level for approximately 150 years.

Different wind fields may be relevant for storm surges at these different locations. For example, van den Brink H. (2020) showed that high storm surges in Harlingen are mainly caused by large-scale depressions with a strong west or northwest component, with a core around or above 60ºN. In contrast, in Hoek van Holland, there is a much greater chance that smaller, more localised depressions generate extreme storm surges. One possible explanation is that the response time of the storm surge in Harlingen may be longer than that in Hoek van Holland, meaning that smaller, faster-moving systems will more frequently lead to extreme water levels in Hoek van Holland as compared to Harlingen. Another possible explanation is that prolonged storms lead to a filling of the North Sea basin, resulting in extra high water levels; a phenomenon which does not occur in Hoek van Holland because the water drains away through the English Channel.

3
Relevant datasets

This chapter provides an overview of the meteorological and hydrodynamic datasets used in this study. The datasets can be divided into two main categories:

1. Datasets used for storm classification and surge analysis in this study:
   a. ***SEAS5*** - Provides mean sea level pressure (MSLP) fields, which are used to classify large-scale atmospheric conditions.
   b. ***DCSM5*** - Provides water level and tidal data at several locations along the Dutch coast, corresponding to the classified weather patterns.
2. Datasets used by (Neal, Fereday, Crocker, & Comer, 2016) to generate and apply their predefined weather patterns.
   a. ***EMULATE*** - Used to generate the predefined weather patterns covering the period 1850–2003.
   b. ***ERA5/ERA-Interim*** - Used to apply the predefined weather patterns to atmospheric conditions from 2004–2022.

3.1 SEAS5, HARMONIE and DCSM5

ECMWF SEAS5

The European Centre for Medium-Range Weather Forecasts System 5 (ECMWF SEAS5) is a seasonal forecasting system developed by ECMWF to provide long-term weather and climate predictions. Launched in November 2017, SEAS5 builds upon previous ECMWF forecasting models, improving predictive capabilities through ensemble forecasting techniques (ECMWF, 2018).

SEAS5 is designed to capture the effects of large-scale climate variations, which can persist for several months. While traditional weather models demonstrate predictive skill up to two weeks ahead, uncertainty increases significantly beyond this period due to the chaotic nature of the atmosphere (ECMWF, 2021). SEAS5 attempts to capture this uncertainty using ensemble forecasting, where multiple simulations are run with slightly varied initial conditions. This approach provides a range of possible outcomes, rather than a single deterministic forecast, offering probabilistic insights into future weather trends.

This concept is illustrated in Figure 3-1, which demonstrates how an ensemble prediction system works. Initially, the forecasts start from nearly identical conditions, but over time, small variations in initial conditions cause the forecast trajectories to diverge. The spread of these trajectories represents the range of possible weather scenarios, with higher uncertainty at longer lead times.

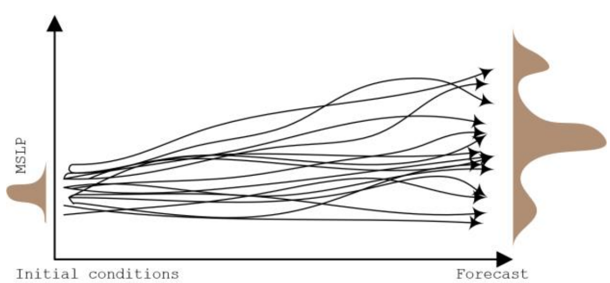


Figure 3-1: Ensemble prediction system as used by ECMWF (adapted from (Grönquist, et al., 2019))

SEAS5 takes various inputs such as historical weather data, sea surface temperatures, and atmospheric composition data to initialise its simulations. It then uses numerical algorithms to simulate atmospheric dynamics, including interactions between the atmosphere and ocean, and predicts future weather conditions. The model produces output variables such as temperature, precipitation, wind speed, and atmospheric pressure (ECMWF, 2021).

Each month, SEAS5 generates an ensemble of 51 global seasonal forecasts, with a lead time of up to seven months. Additionally, hindcasts (reforecasts) spanning 1981–2016 are used to calibrate the system and verify its predictive performance (ECMWF, 2021). SEAS5 has a spatial resolution of 0.25 degrees (approximately 25 km) and a temporal resolution of 6 hours, providing four time steps per day (00:00, 06:00, 12:00, and 18:00).

DCSM5
DCSM5 (Dutch Continental Shelf Model version 5) is a hydrodynamic model developed by Rijkswaterstaat, Deltares, and the Royal Netherlands Meteorological Institute (KNMI). This model is specifically tailored to simulate storm surge dynamics in the Dutch continental shelf region, considering the complex interactions between ocean currents, tides, and atmospheric forcing. The model operates on a 1/12° × 1/18° grid (approximately 8 km × 8 km) over the northwest European shelf region, utilising inputs such as meteorological data, oceanographic observations, and coastal topography to initialise its simulations (van den Brink H. , 2020; Sterl, van den Brink, de Vries, Haarsma, & van Meijgaard, 2009).

To infer surge, DCSM5 utilises meteorological inputs such as mean sea level pressure and wind stress, rather than direct wind speed. It then employs advanced numerical methods to solve the shallow water equations, accounting for factors such as bottom friction, wind stress, and astronomical tide at the open boundaries (van den Brink H. , 2020). The model produces outputs such as water levels, currents, and wave heights.

This model is operationally used by KNMI to predict water levels along the Dutch coast and by Rijkswaterstaat to produce water level forecasts for the North Sea and coastal stations, providing predictions up to 48 hours in advance (Zijderveld, Verboeket, Bosma, & IJpelaar, 2022). It should be noted that the Water Management Centre Netherlands (WMCN) operationally uses DCSM6 for its forecasts. DCSM6 is the successor of DCSM5 and operates at a higher resolution of 1/40° × 1/60° (approximately 1.6 km × 1.6 km), which is five times higher than the resolution of DCSM5 (van den Brink H. , 2020). However, DCSM5 is used for this study, as it was used to process SEAS5 simulations due to its significantly faster computation speed compared to DCSM6.

Furthermore, DCSM5 is integrated with inputs from the HARMONIE model, providing additional meteorological data, and with the ECMWF ensemble model, extending the forecast horizon to up to 10 days (Zijderveld, Verboeket, Bosma, & IJpelaar, 2022).

3.2 EMULATE

As detailed by Ansell et al. (2006), the European and North Atlantic daily to Multidecadal climate variability (EMULATE) dataset was developed to examine atmospheric circulation patterns across Europe and the North Atlantic on a multidecadal scale. Funded by the European Community, the EMULATE project sought to improve understanding of climate variability by creating a long-term, daily mean sea level pressure (MSLP) dataset covering 1850 to 2003. Led by Professor Phil Jones at the University of East Anglia's Climate Research Unit, this project was a collaborative effort involving multiple European institutions, including the Met Office Hadley Centre and other universities across Europe.

Ansell et al. (2006) explain that the primary aim of the EMULATE project was to define and characterise long-term atmospheric circulation patterns by examining trends in persistence, transitions, and the amplitude of dominant patterns over a 154-year period. The project addressed the need for an extended, gridded MSLP dataset, as previous records were constrained by both length and spatial coverage. As a result, EMULATE offers a valuable resource for assessing atmospheric influences on climate extremes and exploring the relationship between circulation patterns and sea surface temperature (SST) trends over the North Atlantic.

The EMULATE dataset's EMSLP fields were generated by blending data from 86 land-based and island pressure stations with marine observations from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) over a 5° latitude-longitude grid. The dataset provides reliable coverage for the North Atlantic-European region (from 70°W to 50°E and 25°N to 70°N). Early EMSLP data (1850-1880) rely solely on sparse terrestrial and marine observations, and from 1881 onwards, the dataset incorporates additional historical Northern Hemisphere gridded MSLP data, enhancing its temporal and spatial coverage.

The EMULATE dataset uses reduced-space optimal interpolation (RSOI) to achieve spatial continuity, allowing for more complete representation in data-sparse regions. This approach enables EMSLP to capture approximately 80-90% of the daily variability when validated against historical and modern reanalysis datasets like ERA-40. Despite its comprehensive design, Ansell et al. (2006) note some limitations in data-sparse areas, such as Greenland and parts of the Middle East, due to limited observational coverage in the early record.

3.3 ERA5/ERA-Interim

The ERA5 and ERA-Interim reanalysis datasets, produced by ECMWF, provide detailed climate and weather information, capturing a wide range of atmospheric, land surface, and oceanographic variables over extended historical periods. These datasets support climate research, model validation, and decision-making for applications like flood prediction, agricultural planning, and renewable energy production.

ERA-Interim

ERA-Interim, spanning from 1979 to August 2019, was ECMWF's primary reanalysis dataset until it was succeeded by ERA5. This dataset has a spatial resolution of approximately 79 km globally and includes 60 vertical levels up to 0.1 hPa. ERA-Interim uses a 4D-Var data assimilation system based on ECMWF's IFS Cycle 31r2 model version, which integrates observational data to provide improved atmospheric state estimates. Data are provided in six-hourly increments, with sub-daily forecast fields available at three-hour intervals up to 24 hours (ECMWF, 2023).

ERA5

ERA5, the successor to ERA-Interim, offers several advancements, covering the period from January 1940 to the present with a significantly enhanced spatial resolution of about 31 km and 137 vertical levels up to 0.01 hPa. Produced using an updated version of ECMWF's IFS (Cycle 41r2), ERA5 provides hourly data, enabling higher temporal granularity. The dataset also includes ensemble data assimilation (EDA) with a ten-member ensemble at 63 km resolution, offering uncertainty estimates to support a wide range of applications in climate monitoring and real-time weather analysis (ECMWF, 2024).

ERA5 replaces ERA-Interim with a higher-resolution, continuously updated dataset that improves on the spatial and temporal detail, precision, and range of applications initially established by ERA-Interim.

PHASE I: APPLYING PREDEFINED WEATHER PATTERNS

Reading Guide

4

Exploratory analysis SEAS5

This chapter provides an exploratory analysis of the SEAS5 dataset to understand its structure and dynamics, focusing on atmospheric drivers of storm surge events along the Dutch coast. The aim is to gain insights into the dataset and identify considerations for its effective use in subsequent analyses.

4.1 SEAS5 dataset overview

The SEAS5 dataset used in this study spans the period from 1981 to 2023, offering extensive temporal coverage of 43 years. This dataset includes nearly 3.5 million daily records, equivalent to over 9,000 years of synthetic data when accounting for the ensemble members. Each day is represented by four time steps: 00:00, 06:00, 12:00, and 18:00. SEAS5 contains a wide range of atmospheric variables, the following of which were included in the dataset provided for this research:

- $mslp$ - Mean sea-level pressure [hPa]
- u_{10} - Zonal wind component at 10 meters [m/s]
- v_{10} - Meridional wind component at 10 meters [m/s]
- $iews$ - Instantaneous eastward turbulent surface stress [N/m²]
- $inss$ - Instantaneous northward turbulent surface stress [N/m²]

Additionally, hydrodynamic variables were computed by KNMI using the Dutch Continental Shelf Model version 5 (DCSM5) based on the SEAS5 dataset. These include:

- $water\ level$ – Total water level [m+NAP]
- $tide$ – predicted tide level [m+NAP]

From these hydrodynamic variables, the storm surge can be inferred as the difference between the water level and the tide.

The SEAS5 dataset is a global model, but for this study, only data within the domain of DCSM5 was used. This dataset covers a geographical area from 13.5°W to 13.5°E longitude and 47.0°N to 65.5°N latitude, encompassing the North Sea, parts of the Dutch coast, and surrounding areas of Western Europe. SEAS5 offers a fine spatial resolution of 0.25 degrees (approximately 25 km), enabling the capture of localised atmospheric and hydrodynamic variations. The spatial domain used in this study, including the grid locations where time series of sea level and tides from DCSM5 were computed, is shown in Figure 4-1.

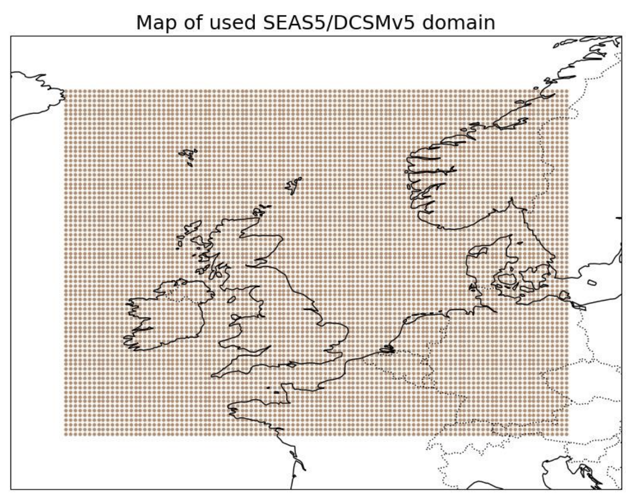


Figure 4-1: Map of the SEAS5 domain used in this study. The grid points indicate the locations where time series of sea level and tide data from DCSMv5 were computed.

4.2 Considerations to account for initial conditions

SEAS5 operates as an ensemble forecasting system, generating 51 global seasonal forecasts every month, each extending up to seven months. To ensure accuracy and reliability in the analysis, only the last six months of each simulation are retained, with the first month discarded. This decision addresses two key challenges associated with initial conditions:

1. ***Strong initial correlation*** **-** The initial conditions for all ensemble members are identical, leading to artificial correlations in the first month. This is evident from Figure 4-2, where the top left panel displays the water level over time, with the mean and ±1 standard deviation highlighted. The top right panel shows the corresponding tide levels, which do not vary across ensemble members as the tide is deterministic. In contrast, the water level does vary across ensembles due to differences in atmospheric conditions. The bottom left panel illustrates the surge values, again with the mean and standard deviation included. The bottom right panel zooms in on the first month, highlighting the initial lack of variability in surge values due to identical initial conditions across ensemble members.
2. ***Initialisation effects*** **-** During the initialisation period, water level calculations begin at zero. This means that, depending on the starting tide level (high or low), the initial water level can create extreme surges (positive or negative) that do not accurately reflect atmospheric influences.

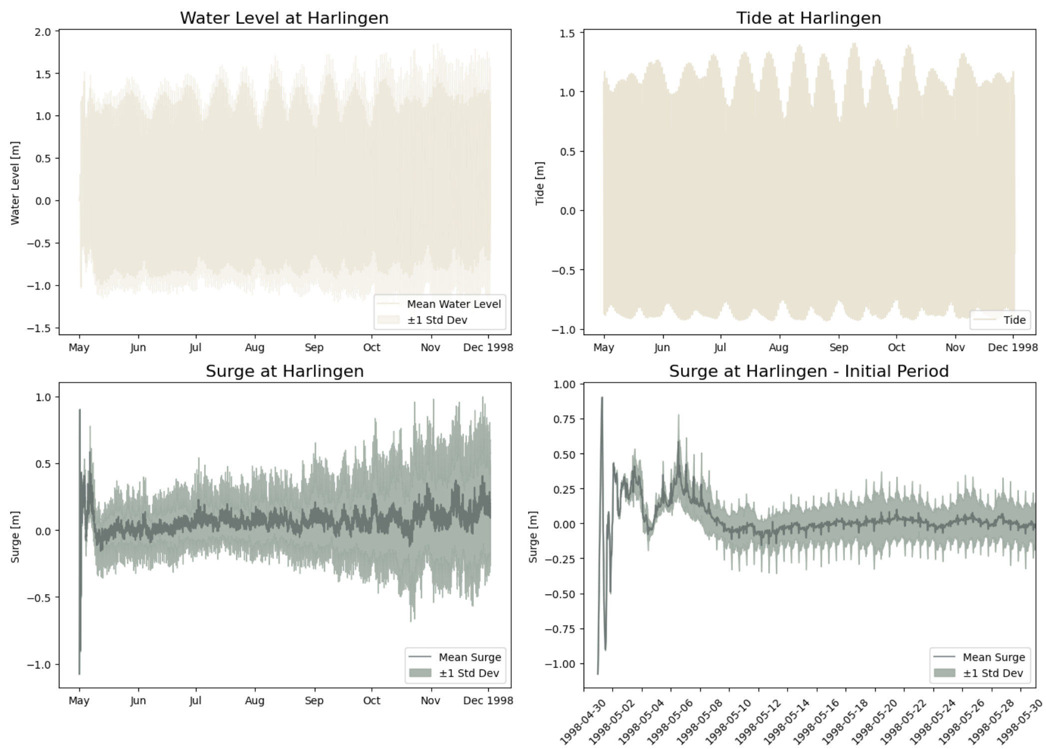


Figure 4-2: Time series showing water level, tide, and surge at Harlingen for a single ensemble member. The initial period is characterised by high variability due to identical initial conditions across SEAS5 ensemble members, leading to artificial correlations. This period is excluded from the analysis.

4.3 Exploration of high and low surge events test

To gain a deeper understanding of the processes driving unusually high or low surge conditions, a case-based exploratory analysis was conducted. By examining representative examples of both high and low surge events, the analysis aims to uncover how atmospheric conditions, such as wind intensity and mean sea level pressure (MSLP), influence surge behaviour along the Dutch coast.

This analysis provides valuable insights into the interplay between atmospheric drivers and observed surge values, offering a more nuanced understanding of the dynamics that govern surge events. By selecting extreme cases, this approach highlights the specific meteorological conditions that correspond to significant high and low surges, helping to illustrate how these conditions manifest in the SEAS5 dataset.

The selected examples are not intended to generalise all surge conditions but rather to serve as a focused, illustrative investigation into the key atmospheric drivers behind extreme coastal surges. This exploration lays the foundation for interpreting surge behaviour in relation to broader-scale weather patterns in later analyses. To investigate, two representative days were selected for closer examination: one with an extreme high surge and another with an extreme low surge, both relative to the entire available SEAS5 dataset.

The following approach was taken for this analysis:

I. The SEAS5 dataset was filtered to identify days with extreme surge, defined as:
   - High surge – A maximum daily surge exceeding $4.5\ m\ +\ NAP$ at any of the six considered locations. A total of 22 such days were identified.
   - Low surge – A maximum daily surge below $-\ 2.0\ m\ +\ NAP$ at any of the six locations. A total of 9 such days were identified.

II. For each identified day, plots of the water level, tide, and surge were generated for various time intervals around the date of interest to confirm its extreme surge.

III. Wind intensity and mean sea level pressure (MSLP) fields were then analysed for each day to confirm the presence of meteorological conditions consistent with the recorded surge values.

Figures 4-3 to 4-5 present both the high surge and a low surge example, illustrating how atmospheric conditions directly influence surge levels. Each of these figures includes two panels, with the high surge example shown on the left and the low surge example on the right.

Figures 4-3 and 4-4 display the extracted maximum daily surge at multiple coastal locations and a close-up of water level, tide, and surge values for one selected site. In both cases, the water level and tide hydrographs align as expected, confirming that these are indeed in-phase. Figures 4-5 and 4-6 further contextualise these surge events by showing surrounding periods: a two-week window and the entire SEAS5 simulation period, respectively. These extended views demonstrate that both the high and low surge events are indeed exceptional within these simulations.

Figure 4-3 presents the MSLP and wind intensity fields from SEAS5 for the selected high and low surge examples. For the high surge example (left panel), the MSLP and wind intensity indicate strong landward winds, which align with the observed high surges. Conversely, the low surge example (right panel) shows strong seaward winds, consistent with the unusually low surge values. Since SEAS5 provides wind stress variables only, the wind speed shown in this figure was derived using the Charnock method, including a bias correction for consistency with the input of the WAQUA-DCSMv5 model. The detailed steps of this conversion process are provided in Appendix B: Theory explained (section B.4).

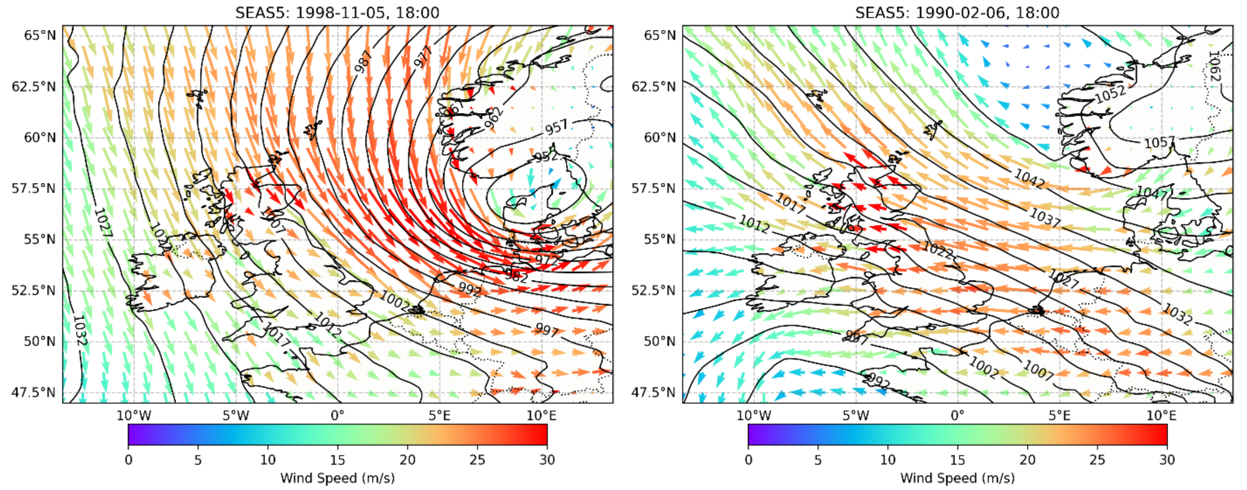


Figure 4-3: SEAS5 MSLP and wind intensity fields for the high surge example (left panel) and low surge example (right panel), illustrating that intense landward wind aligns with high surge and strong seaward wind correlates with low surge.

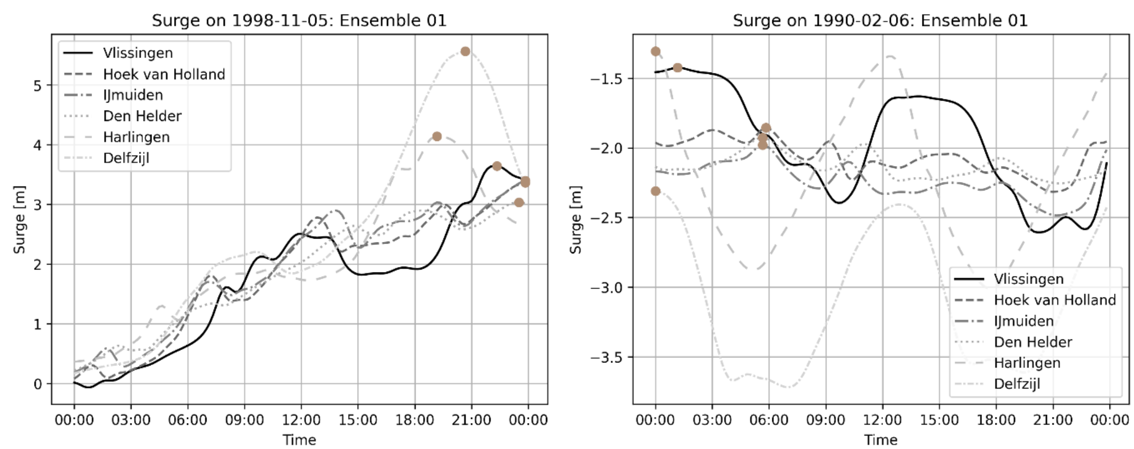


Figure 4-4: Maximum daily surge values at multiple locations for a high surge (left panel) and a low surge example (right panel).

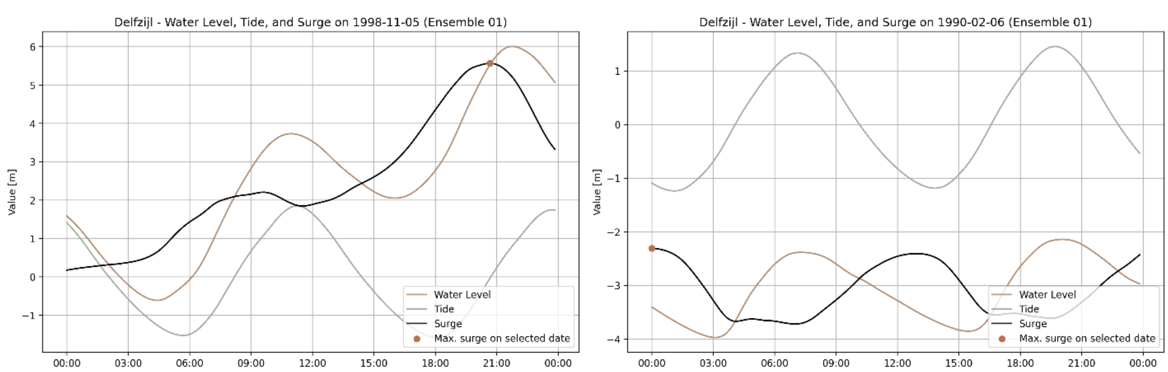


Figure 4-5: Water level, tide, and surge at Delfzijl on the high surge day (left panel) and low surge day (right panel).

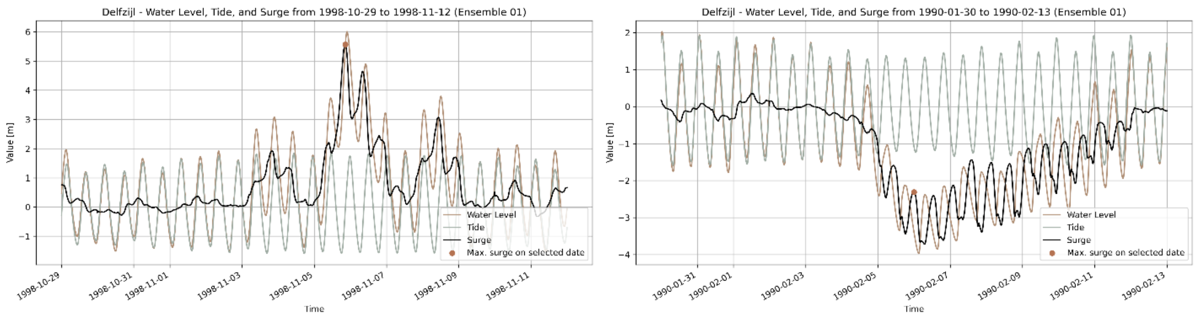


Figure 4-6: Water level, tide, and surge at Delfzijl over a two-week period, centring on the high surge day (left panel) and low surge day (right panel).

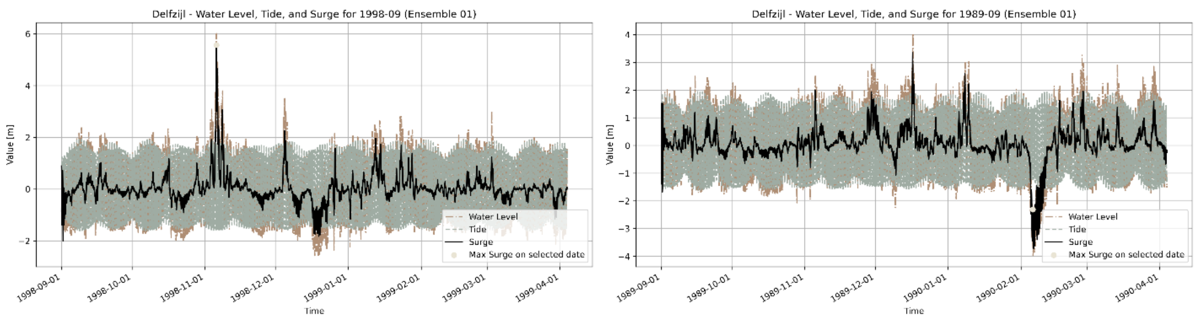


Figure 4-7: Water level, tide, and surge at Delfzijl over the entire SEAS5 simulation period for the high surge (left panel) and low surge (right panel) examples. This overview shows that both surge events represent unusual extremes in the 7-month period.

4.4 Storm tracking within SEAS5 domain

To assess the ability of the SEAS5 dataset to predict extreme surge events in advance, an exploratory analysis was conducted on 22 days with maximum daily surges exceeding 4.5 m. For each of these high-surge events, the mean sea level pressure (MSLP) and wind intensity fields were examined 15 days prior to determine when the corresponding storm systems entered the model domain.

The results indicate that, in all cases, the storm responsible for the extreme surge only entered the domain approximately 1–2 days before the high-surge event occurred. This suggests that attempting to predict extreme surge conditions using MSLP fields 15 days in advance may be inherently challenging within the current domain. The absence of storm systems in the domain at longer lead times implies that a larger spatial domain may be required to capture the early development of storms that contribute to extreme surges along the Dutch coast.

However, despite this limitation, the current SEAS5 domain is retained for this study, as it aligns with the DCSM5 computational domain used to derive corresponding water level and tide predictions. Expanding the domain would require a different hydrodynamic model setup, which is beyond the scope of this research.

An example of one of these storms is provided in Appendix I, illustrating the moment the storm enters the domain and demonstrating the typical evolution of such extreme events.

5
Research method phase I

The Neal weather patterns provide a set of predefined atmospheric configurations designed to classify recurring large-scale weather phenomena. This chapter details the methodology employed in Phase I of this study to match SEAS5 atmospheric fields with Neal's patterns and explore their relationship with storm surge events along the Dutch coast. The aim is to assess the suitability of Neal's predefined patterns for mid-term surge prediction.

The analysis begins with data preprocessing, which aligns the SEAS5 and Neal's 30 weather patterns to ensure consistent spatial domains and resolutions. Following this, SEAS5 atmospheric fields are matched with Neal's patterns using various metrics to identify the best alignment approach. Maximum daily surge values are then extracted and linked to the matched Neal patterns, enabling an investigation of how specific weather patterns contribute to extreme surge events. Finally, surge distributions across the patterns are visualised and evaluated to interpret their relevance for coastal surge prediction. A flowchart which visually shows the steps involved in phase I, may be seen in Figure 5-1.

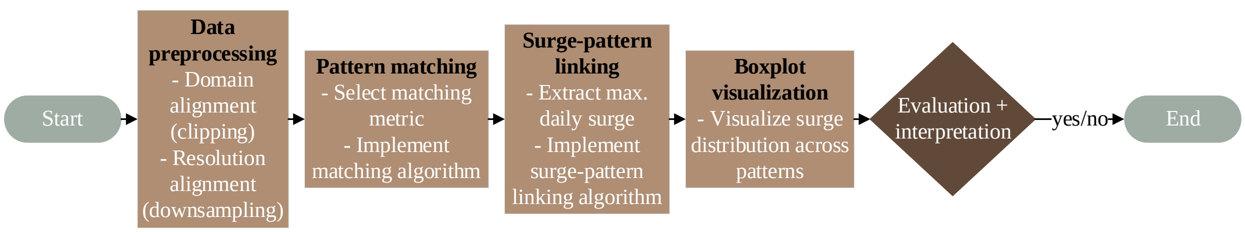


Figure 5-1: Flowchart illustrating the steps involved in phase I.

5.1 Data preprocessing

The first step in comparing the SEAS5 mean sea level pressure (MSLP) fields with the Neal weather patterns involved ensuring a consistent spatial domain and resolution between both datasets.

5.1.1 Domain

As illustrated in Figure 5-2, the SEAS5 and Neal's defined weather patterns cover different geographical extents. The Neal dataset has a broad domain, spanning from 30.0°W to 20.0°E in longitude and from 35.0°N to 70.0°N in latitude, encompassing the UK, parts of Europe, and surrounding seas. This wide range is suited for general weather pattern recognition over Western Europe. In contrast, the used SEAS5 dataset is more geographically focused, covering a smaller area from 13.5°W to 13.5°E longitude and 47.0°N to 65.5°N latitude, which better aligns with the North Sea and the Dutch coast.

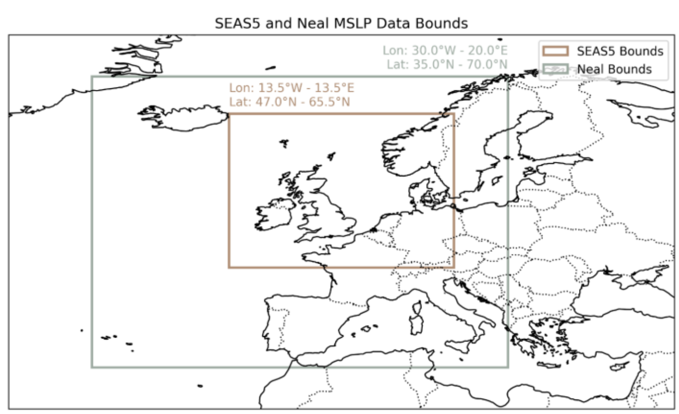


Figure 5-2: SEAS5 and Neal geographic bounds. SEAS5 bounds shown within Neal's broader domain.

This restricted spatial extent of the SEAS5 domain was chosen to match the domain of DCSM5, which was used to compute corresponding water levels and tides based on the SEAS5 atmospheric data.

5.1.2 Resolution

The SEAS5 and Neal datasets also differ in spatial resolution. Neal’s dataset, developed for broader-scale weather pattern recognition, uses a coarser resolution with increments of 5 degrees in both latitude and longitude. SEAS5, optimised for higher-detail meteorological modelling, offers a much finer resolution of 0.25 degrees. This higher resolution captures more localised variations in atmospheric pressure, making it suitable for detailed regional analyses.

5.1.3 Aligning the datasets

Two preprocessing principles were applied to align SEAS5 data with Neal's weather patterns: clipping and downsampling. These steps ensure that both datasets share a common spatial domain and resolution, facilitating accurate comparison.

Clipping

This first step involves adjusting the geographic bounds of the Neal dataset to match the smaller spatial coverage of SEAS5. This adjustment is illustrated in Figure 5-3, which shows the original Neal patterns with SEAS5’s domain marked within black rectangle (left panel), alongside the clipped Neal patterns (right panel). Colour gradients indicate mean sea level pressure, with blue representing lower pressures and red indicating higher pressures.

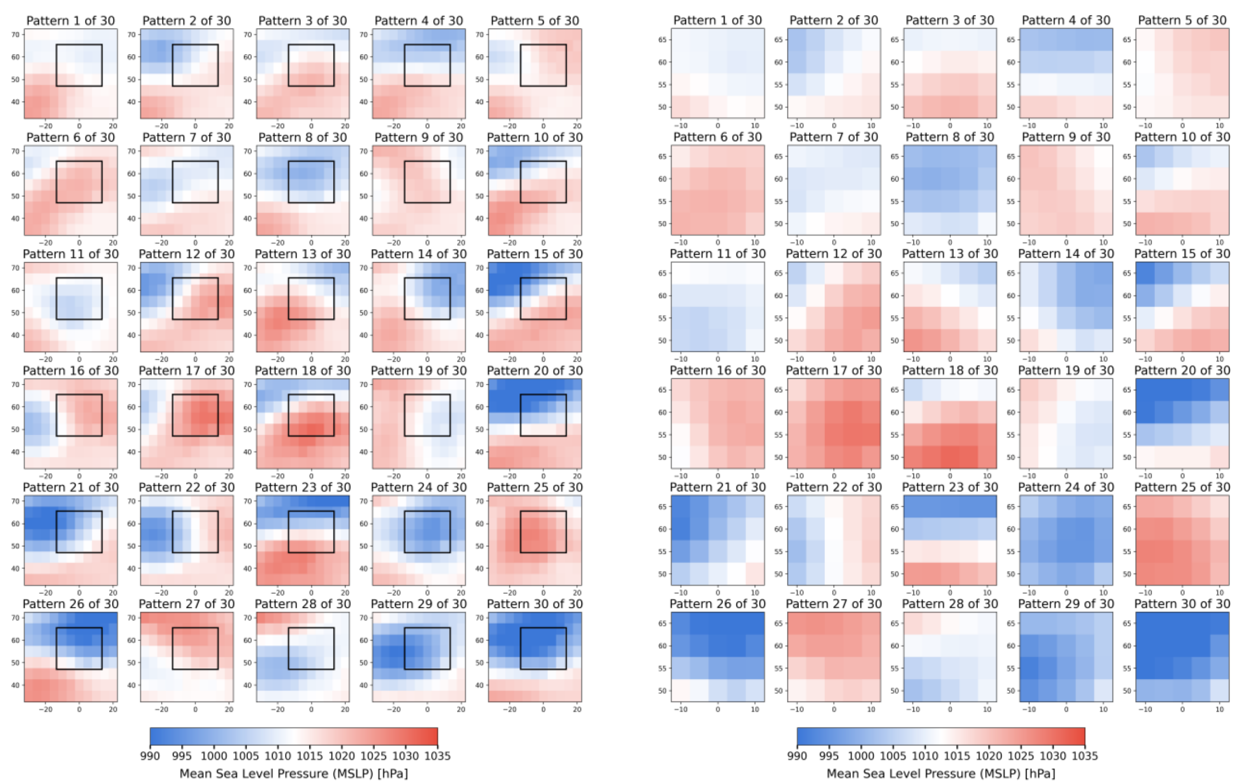


Figure 5-3: Original Neal patterns (left panel) with SEAS5 bounds marked within black rectangle and clipped Neal patterns (right panel). Longitude on x-axis and latitude on y-axis.

Downsampling

Following domain alignment, downsampling is applied to adjust the SEAS5 dataset’s finer resolution to match Neal’s coarser grid.

The downsampling of SEAS5 to match Neal's coarser 5-degree grid involves mapping the high-resolution SEAS5 data (originally at 0.25-degree increments) onto larger grid cells that match Neal's layout and resolution.

1. ***Grid cell alignment -*** SEAS5's finer resolution data points are grouped into blocks that correspond to the larger 5-degree grid cells in Neal. Each 5-degree cell in the Neal grid corresponds to multiple SEAS5 data points due to SEAS5's finer 0.25-degree resolution.
2. ***Nearest neighbour interpolation*** - To downsample, the nearest interpolation method is applied to assign each Neal grid cell the value of the closest SEAS5 data point within that cell. This approach avoids averaging or creating synthetic values, preserving the original SEAS5 information by selecting the nearest observed data point for each coarser grid cell.
3. ***Resampling across entire domain*** - This process is applied across the full SEAS5 domain, producing a downsampled dataset aligned with Neal's grid layout and resolution. The resulting SEAS5 dataset now shares the same spatial resolution and grid points as Neal, allowing consistent comparisons.

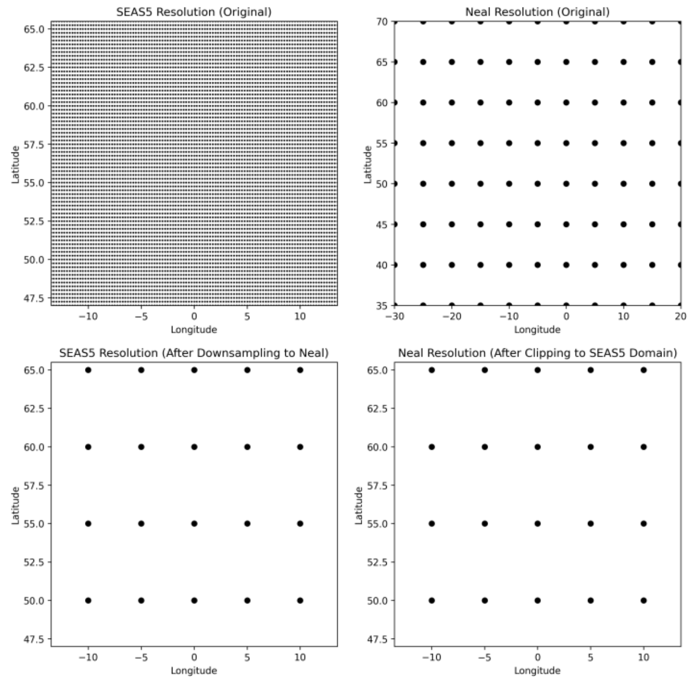


Figure 5-4: Original SEAS5 resolution (top left), original Neal resolution (top right), SEAS5 after downsampling to match Neal's resolution (bottom left), and Neal after clipping to match SEAS5's domain (bottom right).

Downsampling SEAS5 was chosen over upsampling Neal for two main reasons. First, downsampling is computationally more efficient, as it reduces the number of data points without the need for additional interpolation. Second, upsampling Neal through methods like bilinear interpolation would not produce additional meaningful information; instead, it would only estimate intermediate values without improving the inherent data resolution. By aggregating SEAS5 to Neal's grid, a consistent dataset was achieved while preserving computational efficiency and data integrity.

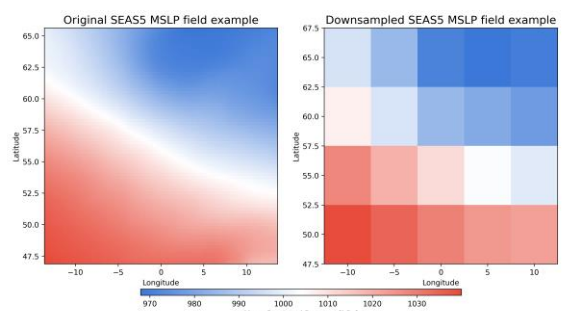


Figure 5-5: Comparison of an original SEAS5 MSLP field (left panel) with its downsampled version (right panel) to match Neal's resolution.

Figure 5-5 presents an example of the downsampling process, comparing an original SEAS5 MSLP field (left) with its downsampled version (right). The down-sampled SEAS5 field shows how localised details captured in the higher-resolution grid are aggregated into broader cells, aligning with the resolution of Neal's patterns.

The result of the combined transformation is illustrated in Figure 5-4, which provides an overview of the clipping and downsampling processes. The top row displays the original datasets: SEAS5 on the left, with its finer resolution and limited domain, and Neal on the right, with its broader domain and coarser grid. The bottom row shows the datasets after preprocessing, with the Neal dataset clipped to SEAS5's domain (bottom right) and SEAS5 downsampled to match Neal's coarser resolution (bottom left). This alignment of domain and resolution establishes a common framework for comparing SEAS5 and Neal data, allowing for the next step: implementing a matching algorithm to link SEAS5 MSLP fields with Neal's predefined weather patterns.

5.2 Matching algorithms

By matching SEAS5 mean sea level pressure (MSLP) fields with the predefined weather patterns from the Neal dataset, it becomes possible to classify SEAS5 atmospheric conditions into specific, recurring patterns that correlate with significant weather phenomena.

Various methods may be used to perform this matching. In this analysis, methods based on Euclidean distance, Root Mean Square Error (RMSE), Pearson correlation and an Autoencoder model are evaluated to determine the most effective way to align SEAS5 data with Neal's patterns.

This section assumes that the reader is familiar with these metrics and their underlying principles. For a detailed explanation of each metric, including their mathematical expressions, please refer to Appendix B: Theory explained (section B.1). The appendix also outlines how these metrics were specifically applied to the SEAS5 and Neal datasets in this study. For further information on autoencoders, including an overview of how they work and the specific architectures tested in this analysis, refer to Appendix A: Autoencoder.

5.3 Linking surge to weather patterns

Following the pattern matching process detailed in the previous section, the next step is to analyse how these matched Neal patterns correspond to surge events along the Dutch coast. The Royal Netherlands Meteorological Institute (KNMI) has computed the corresponding water level and tide data for the synthetic SEAS5 dataset at several locations along the Dutch coast, which is then used to derive surge values. Surge is calculated as the difference between water level and tide, providing a measure of the influence of atmospheric conditions on sea levels.

This analysis is conducted for the locations as previously specified in Section 2.6: Delfzijl, Harlingen, Den Helder, IJmuiden, Hoek van Holland, and Vlissingen. This setup results in a comprehensive dataset that combines atmospheric patterns with surge events. By establishing this surge-pattern relationship, it becomes possible to assess how specific weather patterns contribute to extreme coastal surge conditions, laying the groundwork for improved forecasting of high-risk weather events.

5.3.1 Extracting maximum daily surge

The SEAS5 dataset provides MSLP fields every six hours, while water levels derived from DCSM5 have a time resolution of ten minutes. This results in 144 water level points per day, from which only the maximum daily surge is retained for analysis. Using the full set of 144 water level points for linking would be computationally intensive, and since the focus is on extreme events, using only the maximum daily surge provides a logical and efficient approach. Figure 5-6 illustrates an example of this extraction, highlighting peak surge values for each day (marked in brown).

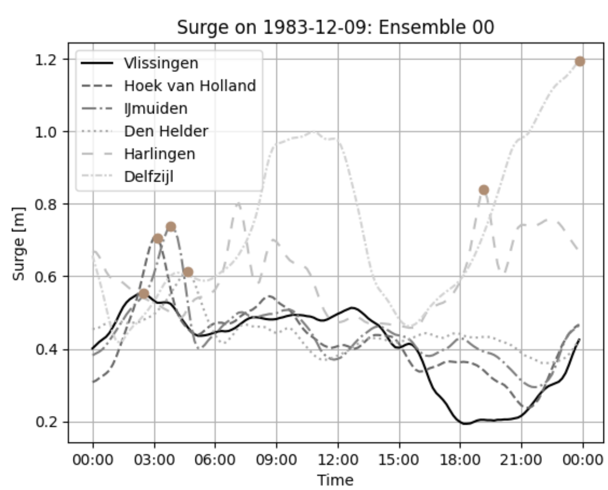


Figure 5-6: Example of maximum daily surge extraction, across multiple locations along the Dutch coast. The brown markers indicate the maximum daily surge values.

5.3.2 Surge-pattern linking

Linking maximum daily surge to Neal's predefined weather patterns is governed by the concept of lead time. Given a specified lead time, the maximum daily surge on a given day is first associated with the corresponding SEAS5 MSLP pattern from the preceding number of days. To establish a connection between SEAS5 data and the broader-scale Neal weather patterns, each SEAS5 MSLP field was first matched to its closest Neal pattern using a predefined similarity metric (see Section 5.2 for the tested similarity metrics). This pre-matching step allows for indirect linking of surge events to Neal patterns through their corresponding SEAS5 atmospheric conditions. This process, illustrated in Figure 5-7, enables the analysis of how prior atmospheric conditions contribute to surge events, supporting mid-term surge forecasting by linking extreme surge occurrences to specific, identifiable weather patterns.

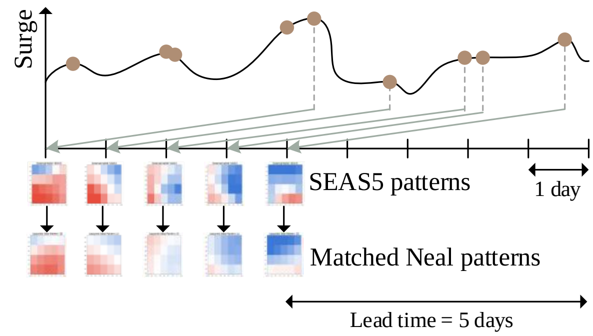


Figure 5-7: Schematic representation of the surge-pattern linking process. The SEAS5 weather pattern of x days before the maximum daily surge (based on the lead time) is assigned to its closest matching Neal pattern..

To account for the lead time, the surge data is adjusted to exclude the initial days for which there is no preceding SEAS5 MSLP pattern to link to. For instance, if a lead time of 5 days is used, surge data starts from the 6th day onward to ensure that every surge event has a corresponding SEAS5 MSLP pattern from the required number of days prior. Since these SEAS5 MSLP patterns have already been assigned to their closest matching Neal pattern, any surge event can be associated with a Neal weather pattern.

Since SEAS5 provides four MSLP fields per day—at 00:00, 06:00, 12:00, and 18:00—the maximum daily surge is linked to the closest preceding SEAS5 MSLP field based on the time of the surge. For example, if a maximum daily surge occurs in the late afternoon, it is linked to the matched SEAS5 pattern at 12:00 rather than the subsequent 18:00 pattern.

5.4 Post-processing

To analyse the relationship between Neal weather patterns and coastal surge events, boxplot visualisations are used to display the distribution of maximum daily surge values associated with each weather pattern. This approach allows for a clear and concise summary of surge variability across the predefined patterns. The primary goal of this visualisation is to identify distinct surge ranges for each weather pattern, facilitating their application in mid-term surge forecasting.

The boxplots are generated by grouping maximum daily surge values based on their corresponding Neal pattern. This process is repeated across multiple locations along the Dutch coast, to capture regional variations in surge behaviour. Additionally, the analysis is conducted for various lead times, reflecting how the influence of atmospheric patterns evolves with forecast horizons.

The ultimate goal for mid-term forecasting, with a target lead time of 15 days, is to derive distinct and well-separated surge ranges for each Neal weather pattern. Such separation enhances the interpretability and utility of the patterns, enabling forecasters to link predicted atmospheric configurations to specific surge levels. This information can then inform decision-making processes, such as scheduling maintenance activities, based on the anticipated impact of forecasted weather patterns on coastal conditions.

6

Evaluation of Neal weather patterns

This chapter presents the results from applying the Neal weather patterns to classify SEAS5 atmospheric data and relate these patterns to coastal surge events. Building on the methodology detailed in Chapter 5, which describes data preprocessing, pattern matching, surge linking and post-processing, this chapter focuses on the analysis and outcomes specific to the Neal patterns. First, a comparison of different matching methods (Euclidean distance, RMSE, and Pearson correlation) is provided to determine the most effective approach for aligning SEAS5 fields with Neal patterns. This is followed by an exploration of surge behaviour, including examples of extreme cases and an assessment of surge distributions across the patterns. The chapter concludes with an evaluation of the Neal patterns' suitability for surge prediction.

6.1 Comparison of matching methods

In section 5.2, several methods were presented for matching SEAS5 mean sea level pressure (MSLP) fields to Neal's predefined weather patterns, including Euclidean distance, Root Mean Square Error, Pearson correlation, and an autoencoder-based approach. This section compares these methods, evaluating their effectiveness in identifying the most appropriate Neal pattern for given SEAS5 fields.

To illustrate the differences between these methods, Figure 6-1 presents an example day with SEAS5 MSLP fields sampled at the four time intervals. The top row shows the downsampled SEAS5 MSLP fields, while the subsequent rows display the corresponding closest-matching Neal patterns based on Euclidean distance, RMSE, and Pearson correlation. This example day was chosen to highlight instances where the methods yield different matching patterns, illustrating the importance of choosing a method that best captures the characteristics of the dataset. It should be noted, however, that for many days, all methods identified the same best matching pattern, indicating a general consistency across approaches.

Generally, Euclidean distance appears to perform the best at matching SEAS5 fields to Neal's patterns in terms of capturing both structural resemblance and intensity of pressure values, based on visual inspection of the results. One important justification for using Euclidean distance is its consistency with the clustering approach used by Neal et al. (2018), who employed k-means clustering (which minimises Euclidean distance) to define the weather patterns. This alignment makes Euclidean distance not only effective but also methodologically consistent and theoretically sound for this application.

An additional observation is the tendency for SEAS5 fields to exhibit more intense MSLP values, both highs and lows, compared to Neal's patterns. This discrepancy occasionally results in no Neal pattern perfectly matching the SEAS5 field's intensity, meaning the Euclidean distance metric does not always achieve a "perfect" match in such cases. However, among the available metrics, Euclidean distance remains the most reliable in capturing the structure and approximate intensity of the atmospheric patterns.

In terms of alternative methods, an autoencoder-based approach was also tested with various architectures. While the autoencoder sometimes demonstrated improved matching of structural elements, such as the relative locations of high and low-pressure regions, it often fell short in accurately capturing the intensity of MSLP values. This discrepancy arose because the autoencoder architecture, as implemented, tended to focus more on reconstructing general spatial structures rather than preserving the exact intensity gradients necessary for robust surge prediction.

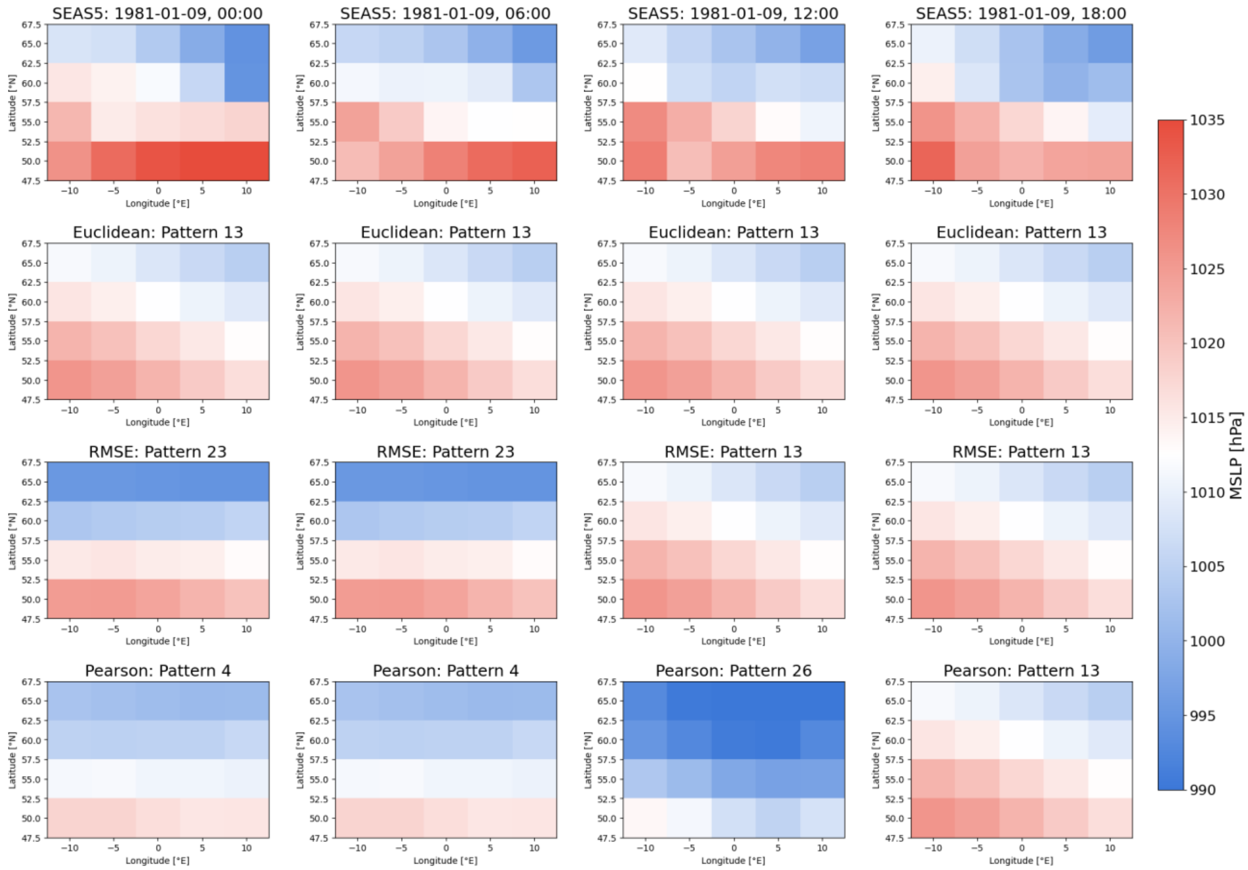


Figure 6-1: Matched SEAS5 MSLP fields (top row) with closest matching Neal patterns based on Euclidean distance (2nd row), RMSE (3rd row) and Pearson correlation (4th row).

Given that Euclidean distance already provides robust results and aligns well with Neal's clustering method, it was ultimately deemed unnecessary to pursue the RGB approach. Consequently, Euclidean distance was selected as the primary matching method due to its effectiveness and consistency with Neal's pattern generation process.

6.2 Comparison of mean, standard deviation and assignment probabilities

The previous section highlighted that SEAS5 often displays more extreme MSLP values, both high and low, than those represented by the Neal patterns. To investigate this observation further, the underlying climatology of SEAS5 was compared with that of the EMULATE dataset, which Neal et al. (2018) used to generate the weather patterns. The EMULATE dataset spans 1850–2003 and SEAS5 covers 1981–2023, making the overlapping period from 1981–2003 a suitable timeframe for direct analysis. This comparison is valuable to determine whether the two datasets share a consistent climatological baseline.

Figure 6-2 illustrates the mean and standard deviation of MSLP for both EMULATE and SEAS5 across the specified period. The right most plots compare the differences between SEAS5 and EMULATE, showing that the mean values are nearly identical, whereas SEAS5 exhibits a notably higher standard deviation. This suggests that, while both datasets capture the same overall climatology in terms of mean MSLP, SEAS5 features a greater variability, with a tendency for more frequent extreme MSLP values. This difference in variability may be explained by differences in how the datasets are generated. EMULATE is a reanalysis dataset constrained by historical observations, meaning its variability is limited to what was actually recorded. SEAS5, in contrast, is a seasonal forecast model, which evolves freely after initialisation, potentially leading to a wider spread of possible MSLP values. Additionally, SEAS5 is an ensemble forecast system, where each ensemble member starts from slightly perturbed initial conditions to account for forecast uncertainty. This ensemble approach could result in a greater range of possible MSLP values, contributing to the observed higher variability in SEAS5.

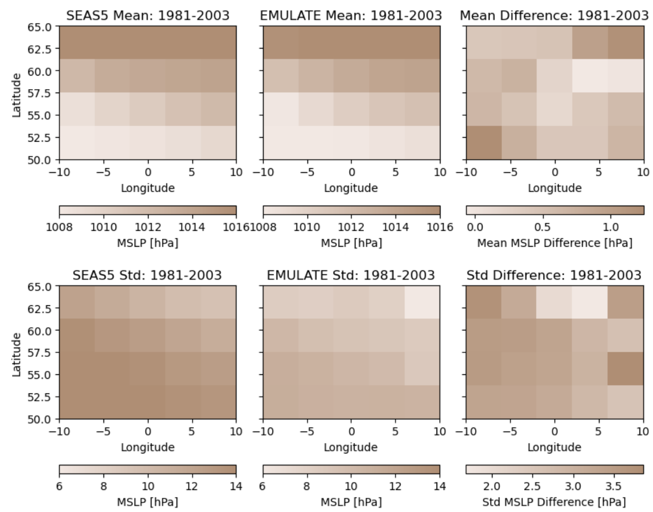


Figure 6-2: Comparison of MSLP mean and standard deviation for SEAS5 (1st column), EMULATE (2nd column), and their differences (SEAS5 - EMULATE, 3rd column) over the 1981–2003 period.

To evaluate the effect of these intensity differences on pattern assignment, the assignment probabilities of each Neal pattern were examined across the two datasets. The results shown in Figure 6-3 indicate discrepancies in how frequently specific patterns are assigned. For instance, SEAS5 assigns Neal pattern 17, which corresponds to the highest MSLP values among the patterns, more frequently than EMULATE. This increased assignment frequency seems to be due to SEAS5's tendency toward more extreme high MSLP values. Figure 6-4 shows an example of this, where the original SEAS5 is shown on the left panel, the downsampled SEAS5 in the middle panel and the assigned Neal pattern on the right panel. While the assigned pattern captures the general structure, SEAS5's MSLP values are noticeably more intense. This effect is observed in the opposite direction as well, where SEAS5 exhibits lower MSLP values than the lowest represented by the Neal patterns. For example, pattern 30, which corresponds to the lowest MSLP values, is assigned more frequently in SEAS5 compared to EMULATE, as seen in Figure 6-3.

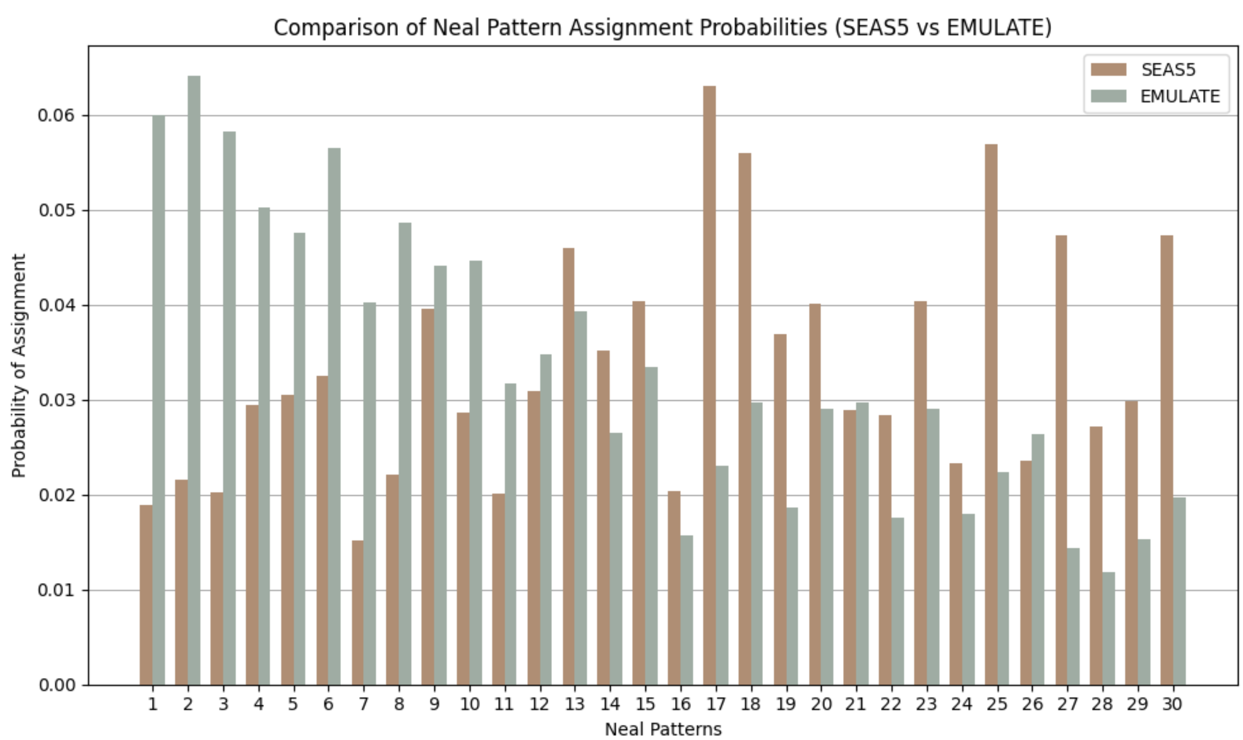


Figure 6-3: Comparison of Neal pattern assignment probabilities for SEAS5 and EMULATE (1981-2003).

These observations suggest that, although SEAS5 and EMULATE are aligned in terms of mean climatology, SEAS5 displays a higher degree of variability, leading to a greater representation of extreme pressure scenarios. This variability likely influences the pattern assignment process, favouring Neal patterns that represent more intense pressure conditions.

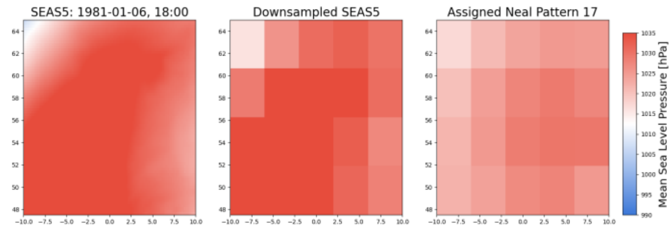


Figure 6-4: Example of SEAS5 assigned to Neal pattern 17.

6.3 Surge distribution across Neal patterns

To investigate the relationship between large-scale weather patterns and coastal surge events, this section presents an analysis of the distribution of daily maximum surge values across Neal weather patterns for various forecast lead times. By examining the surge distribution at different lead times, it is possible to assess whether certain weather patterns are more frequently associated with high surge events, which would indicate predictive value for storm surge forecasting.

To illustrate the distribution of surges across Neal patterns, a series of box plots for IJmuiden is presented, covering lead times of 1, 3, 5, 10, and 15 days (see Figure 6-5). These plots were generated by analysing the entire SEAS5 dataset, with the first month of each simulation excluded to avoid initial ensemble correlation effects, as discussed in Section 4.2. Following the methodology from Section 5.3.2, each SEAS5 time step was matched to the most suitable Neal pattern using Euclidean distance, and daily maximum surge values were computed by subtracting tide from the total water level. Box plots for the other considered locations are provided in Appendix G: Boxplots.

For each lead time, the daily maximum surge values at IJmuiden were grouped according to their assigned Neal patterns, resulting in a box plot that captures the distribution of surge values across the 30 patterns. The boxes represent the 25th, 50th (median) and 75th percentiles of the surge values linked to each pattern, whereas the whiskers extend to the minimum and maximum surge values within each pattern grouping.

The purpose of these box plots is to assess whether certain Neal patterns consistently associate with higher surge values, which could suggest predictive value for storm surge events at various forecast lead times.

Observations

As displayed in Figure 6-5, there is a noticeable trend as lead time increases. At shorter lead times, there is slightly more variability in surge values across some of the patterns. However, as the lead time extends to 15 days, the distributions become more uniformly spread, with no clear concentration of high surges in any particular pattern. This even distribution at longer lead times indicates that high surge events are not strongly associated with specific Neal patterns when forecast lead times are extended.

Additionally, the range of surge values associated with each pattern becomes significantly larger, as indicated by the extended whiskers on the box plots. This wide range further contributes to the limited predictive value of the Neal patterns for anticipating specific surge levels for the SEAS5 dataset, as high or low surges appear across nearly all patterns without clear association.

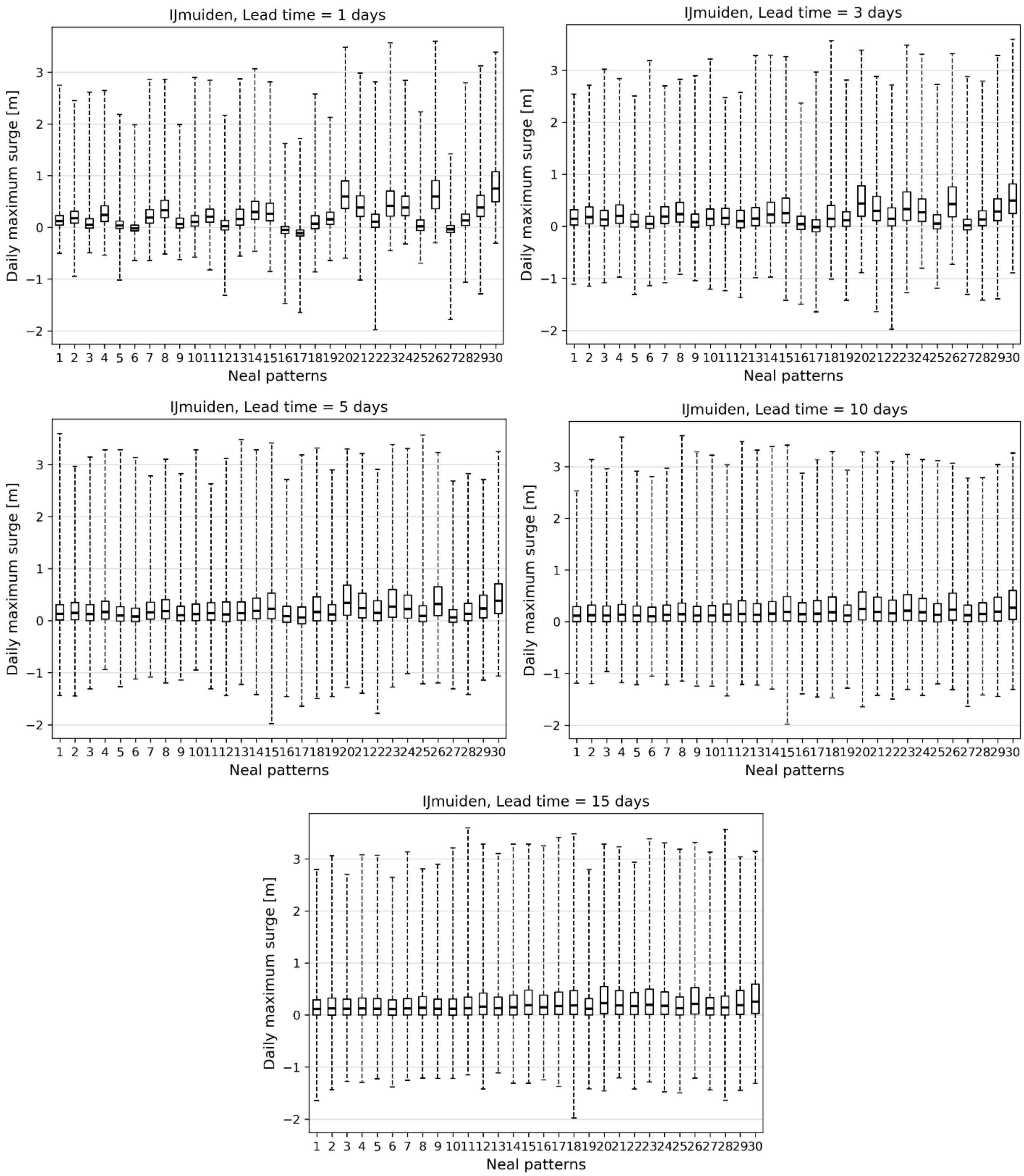


Figure 6-5: Box plots showing the distribution of daily maximum surges at IJmuiden across the 30 Neal weather patterns for lead times of 1, 3, 5, 10, and 15 days. Each box represents the interquartile range (25th to 75th percentiles), with the median shown as a horizontal line, while whiskers indicate the minimum and maximum surge values. These plots illustrate the diminishing association between specific Neal patterns and high surges as the lead time increases.

6.4 Conclusions and next steps

The findings from Phase I highlight several potential limitations in the use of Neal weather patterns for mid-term surge prediction along the Dutch coast. These limitations, as well as potential solutions, are summarised below:

1. ***Matching method selection*** - The comparison of different pattern-matching methods demonstrated that Euclidean distance is the most effective approach for aligning SEAS5 MSLP fields with Neal's predefined weather patterns. This method was selected due to its effectiveness in capturing both structural resemblance and intensity variations, as well as its consistency with the k-means clustering method originally used by Neal et al. (2016) to generate the weather patterns.
2. ***Variability in SEAS5 data*** - The analysis revealed that while SEAS5 and the EMULATE dataset are aligned in terms of mean climatology, SEAS5 exhibits greater variability, particularly in the intensity of high and low MSLP values. This increased variability likely influences the pattern assignment process, favouring Neal patterns that represent more extreme pressure conditions. This discrepancy suggests that Neal patterns, developed using the EMULATE dataset, may not be optimally suited for classifying SEAS5 atmospheric data, as SEAS5 features more frequent extreme MSLP values.
3. ***Domain size considerations*** - The spatial domain used in this study, which aligns with DCSM5, may be too small to capture the early development of storms that influence surge events at longer lead times (e.g., 15 days in advance). An analysis in Section 4.4 showed that for nearly all major storms in the SEAS5 dataset, the storm system only entered the domain approximately 1–2 days before the peak surge occurred. This indicates that the current domain size may not fully capture the atmospheric precursors of extreme surge events when using longer lead times. While this issue is related to the chosen domain rather than the Neal patterns themselves, a potential solution is to expand the spatial domain or implement clustering methods that account for evolving atmospheric conditions over multiple forecast days.
4. ***Suitability of Neal patterns for surge prediction*** - The Neal weather patterns were designed for a broad range of applications, including medium-range probabilistic weather forecasting and assessing coastal flooding risks for the UK. As such, they may not be well-suited for capturing the specific dynamics that drive storm surges along the Dutch coast. One way to address this is to incorporate the goal of surge prediction directly into the clustering process. For example, creating specific clusters tailored to extreme surge events could yield patterns that are more effective for forecasting. The boxplots presented in Figure 6-5 and Appendix G: Boxplots indicate that the Neal weather patterns may not provide a reliable basis for forecasting high surges far in advance using the SEAS5 dataset. These plot showed that at shorter lead times, some Neal patterns exhibit a stronger association with high surge values. However, as the lead time increases to 15 days, the distribution of surge values becomes more uniform across patterns, and the range of associated surge values expands significantly. This suggests that Neal patterns do not strongly differentiate between high and low surge conditions at longer lead times, limiting their predictive value for mid-term surge forecasting. This conclusion holds for all considered locations along the Dutch coast.

Based on these findings, the research will shift focus to alternative pattern classification methods in subsequent chapters. Phase II will explore the creation of new patterns by clustering of the SEAS5 dataset directly. The goal is to develop weather patterns with stronger correlations to surge events, potentially enhancing mid-term range surge prediction capabilities.

PHASE II: CLUSTERING NEW WEATHER PATTERNS

Reading Guide

7

Research method phase II

This chapter details the methodology employed to generate weather patterns using dimensionality reduction via Principal Component Analysis (PCA) and clustering through k-means.

The goal of this methodology is to classify daily atmospheric fields into distinct clusters, with each cluster representing a unique weather regime. The analysis begins with the selection of data from the SEAS5 dataset, focusing on days with extreme surge conditions. In the data preprocessing step, relevant predictor variables are selected. Different atmospheric or hydrodynamic variables could be used depending on the objective and desired complexity. In this study mean sea-level pressure (MSLP) and its gradient are chosen. Dimensionality reduction via PCA is then applied to reduce the complexity of the data while retaining the most significant patterns of variability. Finally, k-means clustering is performed to group similar days into distinct weather patterns. The resulting patterns are analysed and visualised to interpret their meteorological significance. A flowchart which visually shows the steps involved in phase II, may be seen in Figure 7-1. The final step in this process is an evaluation step, where the generated weather patterns are assessed for their suitability in mid-term forecasting of high surge events. If the patterns do not sufficiently capture the relevant atmospheric conditions, adjustments can be made, such as modifying the number of clusters (k), selecting different predictors, or refining the data selection method before re-running the clustering process.

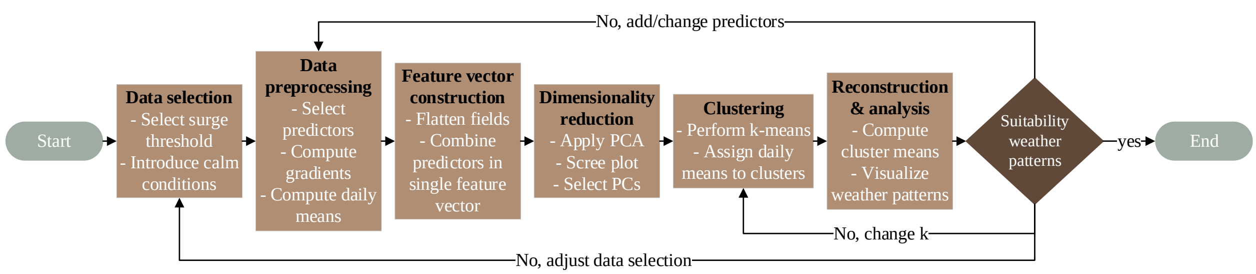


Figure 7-1: Flowchart illustrating the steps involved in phase II.

7.1 Data selection

The SEAS5 dataset used in this study spans the period 1981 to 2023, offering an extensive temporal coverage of 43 years. Each month, SEAS5 provides a seasonal forecast of up to 7 months ahead, with predictions generated for multiple ensemble members. These ensemble members represent slightly perturbed versions of the initial atmospheric conditions, designed to account for uncertainty in the forecast. This comprehensive dataset includes a total of 3,438,035 daily records, corresponding to approximately 9,420 years of synthetic data when combining all ensemble members. Each day is represented by four 6-hourly timestamps.

Given the sheer size of this dataset a targeted sub-selection of the data for clustering is necessary for two main reasons:

I. ***Computational efficiency*** - The first reason is computational efficiency. Due to the large size of the SEAS5 dataset, working with the full dataset would be computationally expensive and time-consuming. A sub-selection is necessary to ensure the analysis is manageable within the computational constraints.

II. ***Focus on extreme events*** - The second reason for sub-selection is related to the focus of the study: forecasting extreme weather events. By definition, extreme surge events occur infrequently. If clustering were performed on the entire SEAS5 dataset, the resulting weather patterns would primarily capture common, calmer conditions, making it difficult to distinguish the atmospheric patterns responsible for extreme surges. By selecting more days with extreme surge conditions and fewer days with calm conditions, the clustering process is optimised to better represent the atmospheric conditions linked to high surges.

There are many ways to perform such a targeted sub-selection, and this study explores three different approaches: fixed surge threshold, Maximum Dissimilarity Algorithm (MDA), and stratified sampling approaches, which are described more fully in Chapter 8.

To illustrate one of these methods, Figure 7-2 shows the surge threshold approach. In this method, all days where the maximum daily surge exceeds a predefined threshold are selected for clustering. This ensures that the identified weather patterns reflect the atmospheric conditions associated with extreme surges. Once the weather patterns are generated, the analysis follows a similar methodology to Phase I. Specifically, the maximum daily surge values from the entire SEAS5 dataset are linked to their closest matching weather pattern, now referred to as Najda patterns, using a predefined lead time. This step is conceptually identical to the approach used in Phase I, where Neal patterns were assigned based on SEAS5 MSLP fields. This process was visually depicted in Figure 5-7 and the reader may refer to it while mentally substituting Neal patterns with Najda patterns in this context.

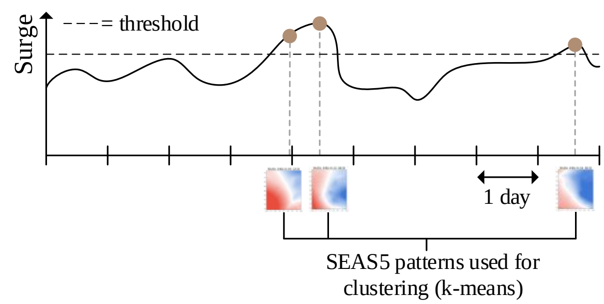


Figure 7-2: Schematic representation of the data selection process for clustering. In this example, all days where the maximum daily surge exceeds a predefined threshold are selected for clustering.

7.2 Data preprocessing

The next step is to preprocess the selected SEAS5 dataset for clustering. This process involves selecting relevant atmospheric variables, computing gradients, and averaging temporal data while reducing noise and dimensionality. The resulting pre-processed data forms the foundation for constructing feature vectors and subsequent analysis.

7.2.1 Selection of predictors

The first step involves selecting the atmospheric variables that will serve as predictors for clustering. For this study, mean sea-level pressure (MSLP) and its squared gradient are chosen. MSLP provides information about the distribution of atmospheric pressure, which is closely tied to large-scale circulation patterns. High-pressure and low-pressure systems, for example, are directly reflected in MSLP fields and play a significant role in driving regional weather.

The squared gradient of MSLP is included to capture spatial variations in pressure. This gradient represents the rate of pressure change across a given distance, which is closely associated with wind intensity and storm-related dynamics. The spacing of isobars determines wind speed, with tighter isobar spacing indicating stronger winds, while their orientation provides insight into wind direction (Hegermiller, et al., 2017). However, wind does not flow directly from high to low pressure, as the Coriolis effect deflects air movement to the right in the Northern Hemisphere, causing winds to flow parallel to isobars in a geostrophic balance (Holton & Hakim, 2013). In reality, surface friction further modifies wind trajectories, causing winds to cross isobars at an angle toward lower pressure. Over land, friction is greater, leading to a more pronounced deviation from geostrophic flow, while over the ocean, winds align more closely with geostrophic balance (Wai-hung, 2010).

Including multiple predictors can enhance the clustering process by offering a richer representation of atmospheric conditions, but it also requires careful handling of two important aspects:

I. Managing data dimensionality, which is addressed through dimensionality reduction techniques such as Principal Component Analysis (PCA).
II. Ensuring comparability between predictors, as differences in magnitude can cause variables with larger absolute values to dominate the clustering process. To prevent this, standardisation or normalisation techniques are applied to scale all predictors to a comparable range before clustering.

7.2.2 Computation of gradients

The gradient of MSLP is computed using central finite differencing, a numerical technique that estimates spatial changes between neighbouring grid points. While this method ensures accurate representations of pressure gradients, it requires excluding the outermost rows and columns of the domain, as these points lack sufficient neighbours for gradient calculations. Consequently, the original spatial grid of 75×109 is reduced to 73×105, which defines the domain for further analysis. To maintain consistency, the MSLP fields are also restricted to this reduced domain.

7.2.3 Temporal averaging

To focus on synoptic-scale variability, daily averages of the six-hourly SEAS5 fields are computed. Daily averaging filters out high-frequency fluctuations, such as diurnal variability, while preserving the larger-scale patterns that are critical for clustering. This process produces a dataset of daily mean fields for both MSLP and its gradient, which form the basis for feature extraction in subsequent steps.

7.3 Feature vector construction

After preprocessing, the daily atmospheric fields are transformed into feature vectors. A feature vector represents the spatial configuration of the selected atmospheric variables for a single day. For each variable, the two-dimensional grid (e.g., 73×105) is flattened into a one-dimensional vector. For example, an MSLP field with 73×105 grid points is reshaped into a vector of size 7,665. The gradient fields undergo the same transformation. These flattened vectors are then concatenated to form a single feature vector for each day. For two predictors (MSLP and its gradient), the resulting feature vector has a size of 15,330 (2×7,665).

The daily feature vectors are combined into a standardised matrix. Each row of this matrix corresponds to a single day, and each column represents one of the 15,330 features. This feature matrix serves as the input for dimensionality reduction using PCA, described in the next section.

7.4 Dimensionality reduction using PCA

PCA is applied to address the challenges of high-dimensional data. By projecting the feature vectors onto a lower-dimensional space, PCA retains the dominant patterns of variability while significantly reducing computational complexity.

Before applying PCA, the dataset is standardised to ensure that all predictors contribute equally to the analysis. Since MSLP and its gradient have different magnitudes, standardisation is necessary to prevent variables with larger absolute values from dominating the principal components.

PCA operates by identifying orthogonal axes, called principal components, that capture the largest variance in the dataset. These components allow for dimensionality reduction and noise filtering, enabling clustering algorithms like k-means to perform more effectively (Jaadi, 2024). A more detailed explanation of how PCA works may be found in Appendix B: Theory explained (Section B.3). This includes both a visual and mathematical explanation, as well as more background on the reasoning for applying PCA in this study.

The number of components retained is determined by analysing the cumulative explained variance, which quantifies how much of the total variability is captured by the principal components. In this study, components are retained until they explain 99% of the variance. Figure 7-3 presents an example of a cumulative scree plot (based on the fixed surge threshold method), illustrating this selection process. Note that this chapter focuses on explaining the general methodology and that the results for other data selection methods are provided in the next chapters.

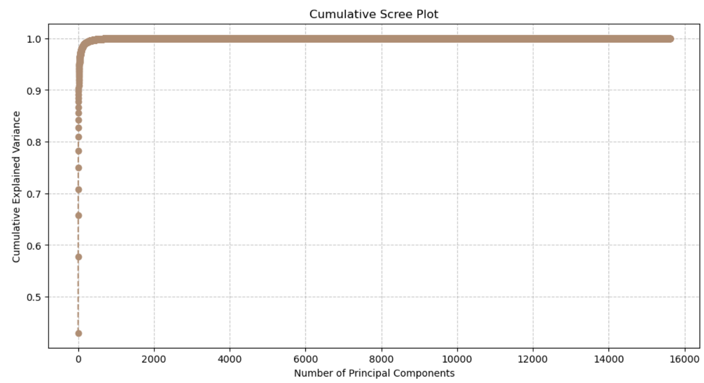


Figure 7-3: Cumulative explained variance as a function of the number of principal components.

This significantly reduces the dimensionality of the data, transforming the original feature matrix into a latent space. In this new space, each principal component is a linear combination of the original features (i.e., MSLP and its gradient at all grid points). The reduced feature matrix, which still contains the same number of rows (representing days) as the original matrix, now has far fewer columns, making it computationally efficient and well-suited for clustering while still retaining the dominant patterns of variability in the data.

7.5 Clustering using k-means

The reduced feature matrix obtained after PCA is used as the input for k-means clustering. K-means is an unsupervised machine learning algorithm that partitions a dataset into k clusters by iteratively minimising intra-cluster variance and maximising inter-cluster separation. For a detailed explanation of the k-means algorithm, including its mathematical foundation and iterative process, refer to Appendix B: Theory explained (Section B.2). A simplified visual example is also presented in this appendix.

The choice of k, the number of clusters, affects the granularity of the resulting weather patterns. Larger values of k yield more specific patterns, which capture finer-scale variations, while smaller values produce broader regimes that encompass general atmospheric trends. The optimal value of k can be determined through exploratory analysis and various validation metrics. This study uses two such metrics: Within-Cluster Sum of Squares (WCSS) and Between-Cluster Sum of Squares (BCSS), to evaluate the cohesion and separation of the clusters, respectively. WCSS measures how tightly data points are grouped within each cluster, while BCSS quantifies the separation between clusters. For a mathematical description as well as a visual example of these metrics, see Appendix B: Theory explained (Section B.2).

Figure 7-4 shows an example of the behaviour of WCSS and BCSS as functions of k for the fixed surge threshold method. As k increases, WCSS decreases because the clusters become smaller and more cohesive, while BCSS initially increases as the clusters become more distinct, eventually stabilising. Using the elbow method, $k = 40$ is selected as the optimal value for clustering in this case. This choice balances the trade-off between minimising WCSS (cohesion) and maximising BCSS (separation), as further increases in k yield diminishing returns in both metrics.

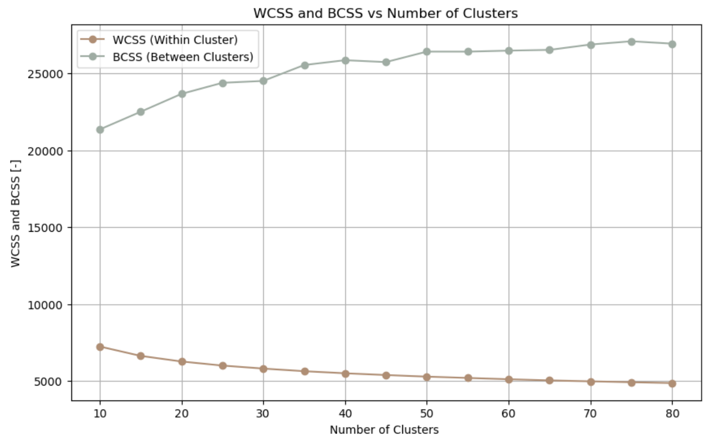


Figure 7-4: WCSS and BCSS as functions of the number of clusters (k). Here, $k = 40$ is chosen to be optimal using the elbow method.

The k-means algorithm yields two outputs: the cluster centroids and the cluster assignments for each day. The centroids represent the mean feature vector for each cluster, providing a numerical summary of the corresponding weather pattern in the reduced latent space. The cluster assignments associate each day with a specific centroid, effectively grouping the days into distinct weather regimes. These outputs form the foundation for further analysis and visualisation of the characteristic weather patterns in the next sections.

7.6 Analysis and visualisation of weather patterns

After clustering, the mean fields for each cluster are reconstructed to represent the characteristic weather patterns. For each cluster, the mean MSLP and gradient fields are computed by averaging all daily fields assigned to that cluster. These mean fields provide a clear and interpretable representation of the spatial configurations associated with each weather regime.

Since the clustering is performed in the reduced dimensional space of principal components (PCs), it is necessary to transform the cluster centroids back into the original feature space to reconstruct the weather patterns. Principal components are linear combinations of the original features, so this transformation involves reversing the PCA projection by applying the principal component weights to the cluster centroids. This step ensures that the reconstructed weather patterns are expressed in terms of the original variables (MSLP and its gradient) on their full spatial grid. This process effectively bridges the reduced latent space of the PCA with the physical meteorological fields.

The weather patterns are generated by:

1. ***Assigning daily fields*** - Each day is assigned to the cluster whose centroid it is closest to in the reduced feature space.
2. ***Averaging assigned fields*** - For each cluster, the daily MSLP and gradient fields associated with the cluster are averaged, producing mean MSLP and gradient fields.
3. ***Transforming centroids back*** - The cluster centroids are projected back from the reduced space to the original feature space to reconstruct the spatial representation of the characteristic weather patterns.

The reconstructed patterns are visualised as spatial maps, which enable the identification of key features such as high-pressure ridges, low-pressure systems, and frontal zones. These patterns can be analysed to understand their meteorological significance and their potential impact on specific weather events.

7.7 Advantages and limitations of the approach

This methodology offers several advantages. PCA reduces computational costs and enhances clustering performance by focusing on dominant atmospheric patterns. K-means clustering is straightforward to implement and provides interpretable results. The possibility of including additional predictors, such as wind stress, offers flexibility to refine the analysis.

However, the approach is not without limitations. The results are highly dependent on the selected predictors, and insufficient predictors may fail to capture critical atmospheric features. Additionally, the choice of k is subjective and may require iterative testing to optimise.

8

Data selection

In this chapter, three different methods for selecting days from the SEAS5 dataset are compared: fixed surge threshold, Maximum Dissimilarity Algorithm (MDA), and stratified sampling. Each of these approaches aims to generate a representative subset of data that captures the full range of surge conditions, but they differ in their methods of selection and the computational cost involved. The fixed surge threshold method (Method A) selects days where the maximum daily surge exceeds a defined threshold, MDA (Method B) focuses on selecting days that maximise dissimilarity in atmospheric conditions, and stratified sampling (Method C) ensures balanced representation across different surge categories.

For all three methods, the selection of SEAS5 days for clustering is based on a surge criterion, though the exact approach differs per method. Once a maximum daily surge event is identified, the corresponding day in SEAS5 is selected, and the daily mean SEAS5 MSLP field for that day is used as input for clustering (see Figure 7-2 for a schematisation of this). These selected atmospheric patterns are then grouped into distinct weather patterns using k-means clustering.

After the weather patterns are generated, the methodology follows the same approach as in Phase I: the entire SEAS5 dataset is linked to its closest-matching weather pattern based on a chosen lead time (see Figure 5-7 for a schematisation of this). This allows for the evaluation of surge distributions across different patterns, which is ultimately visualised using boxplots.

8.1 Method A: Fixed surge threshold

The surge threshold approach for data selection focuses on identifying extreme weather conditions by setting a predefined surge threshold. Specifically, days with a maximum daily surge greater than 1.5 meters at any of the six considered locations along the Dutch coastline were selected. On these selected days, surge values from all six locations are included in the dataset, even if at some locations the surge was below the surge threshold. This threshold was chosen as a representative indicator of "extreme" surge conditions, capturing significant but not overly rare storm events. This threshold of 1.5 meters corresponds to varying return periods depending on the location along the coastline. For example, at Delfzijl, such surge levels may occur multiple times a year, while at Hoek van Holland, they are expected roughly once every seven years. These return periods were determined through an internal analysis conducted by WMCN, the results of which are detailed in Appendix C: Surge return period analysis.

While this may not represent extremely rare events, it is particularly relevant for practical applications such as planning and preparing for maintenance along the Dutch coast, which often occurs on an annual basis. Including these conditions ensures that significant but

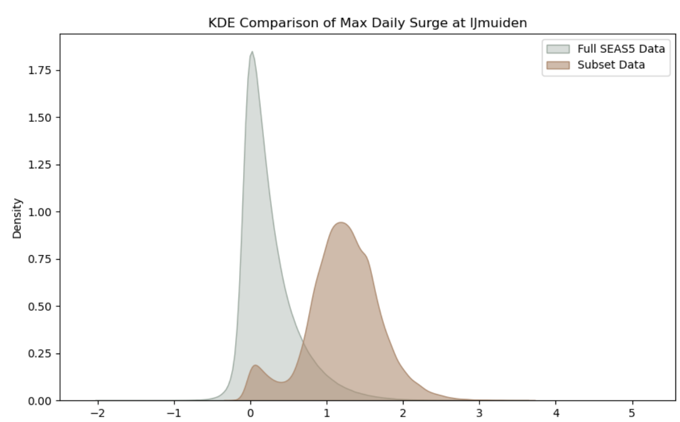


Figure 8-1: KDE plot comparing the full SEAS5 data with the subset selected using a 1.5m surge threshold and additional calm days.

not overly rare storm surge events are captured in the analysis, making the weather patterns actionable for medium-term forecasting and preparation.

If clustering were performed solely on the extreme surge dataset (days with surge > 1.5 [m]) however, the resulting weather patterns would fail to represent calm conditions. This omission would introduce bias when applying the weather patterns to the full SEAS5 dataset for surge predictions. Specifically, days with lower surges would likely be misclassified into inappropriate weather patterns, as no clusters corresponding to such conditions would exist.

To address this limitation, a subset of calm condition days was also included in the clustering. Calm conditions are defined as days where the maximum daily surge falls between 0 and 1.5 meters. To balance the dataset without overwhelming it with calm conditions, the number of calm days added was set to 10% of the extreme surge days. This brings the total number of days used for clustering to approximately 123,750 days, which constitutes about 3.6% of the entire SEAS5 dataset (3,438,035 days).

Figure 8-1 illustrates the distribution of the full SEAS5 dataset compared to the selected subset. As shown, the subset data is more focused on the higher surges, as indicated by the larger right-most peak, while the smaller peak represents the 10% calm conditions included in the dataset. Additionally, because all locations are included whenever the threshold is exceeded at any one location, many surge values below 1.5 meters remain in the selection. This explains why the selected dataset contains a range of lower surge values, beyond those added explicitly as calm conditions. This balanced approach ensures that the resulting weather patterns account for both extreme weather conditions and typical, calmer conditions, allowing for more accurate classification and interpretation of the full SEAS5 dataset.

8.2 Method B: Maximum Dissimilarity Algorithm

The Maximum Dissimilarity Algorithm (MDA) (Camus, Mendez, Medina, & Cofiño, 2011) was chosen to explore how the selection of days could be optimised by maximising the dissimilarity between them. The core idea of MDA is to start by selecting a single day, then iteratively selecting days that are maximally different from the already selected days in terms of their surge values. For a more detailed explanation of the theory behind MDA, including a visual example, refer to Appendix B: Theory explained (section B.5).

Initially, the goal was to apply the MDA approach across all six surge locations. However, running MDA on the full SEAS5 dataset proved to be computationally intensive and time-consuming, making it impractical. To address this, IJmuiden was chosen as the sole location for applying MDA, as it is a central location, making it a suitable representative of the overall surge conditions along the Dutch coast. To further reduce the computational cost, a random sample of 10,000 days was first selected, representing approximately 0.29% of the total SEAS5 dataset. This random selection was done to capture a broad distribution that mirrors the full SEAS5 dataset, as the distribution of the 10k random sample was found to be very similar to that of the full SEAS5 dataset (as shown in the KDE plot in Figure 8-2).

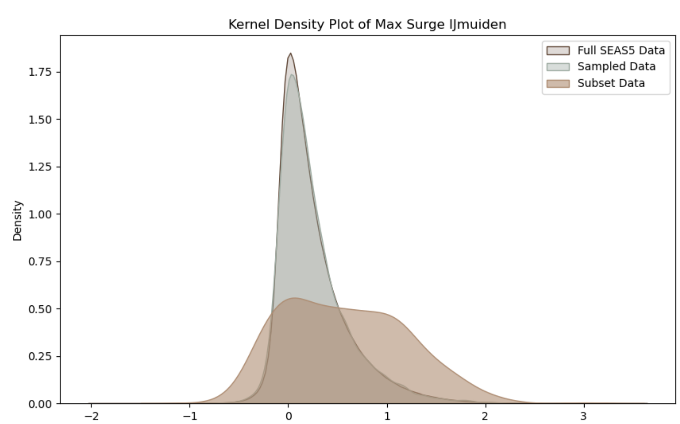


Figure 8-2: KDE plot comparing the full SEAS5 data with the subset selected via Maximum Dissimilarity Algorithm (1,000 days).

Once the random sample of 10,000 days was obtained, MDA was applied to this smaller subset. However, even with this reduced dataset, the computational cost of MDA was still high. As a result, the final number of 1,000 selected days was chosen. The first day was chosen as the day with the highest surge at IJmuiden from the 10,000-day random sample. Subsequent days were selected based on their dissimilarity to the already selected days. By maximising the dissimilarity between selected days, MDA guarantees that even the extreme surge conditions in the tails of the distribution are sufficiently represented, which would not have been the case with random sampling, where surge density would have influenced the selection.

8.3 Method C: Stratified sampling

The stratified sampling approach (Parsons, 2017) was introduced as a response to the computational limitations encountered with the MDA. Unlike MDA, which aims to select days that are maximally dissimilar from each other, stratified sampling divides the data into distinct groups (or "strata") based on surge levels and ensures that each group is represented. For a more theoretical explanation of stratified sampling, refer to Appendix B: Theory explained (section B.6).

For the stratified sampling approach, the SEAS5 dataset was divided into nine bins, each representing a 0.5-meter difference in surge height. The bins range from negative to positive surge values, allowing for a nuanced classification of different surge conditions. The surge categories were defined as follows:

- The lowest bin includes surge values less than -2 meters.
- The highest bin includes surge values greater than 3 meters.
- The intermediate bins covers surge values between -2 and 3 meters, with 0.5-meter intervals.

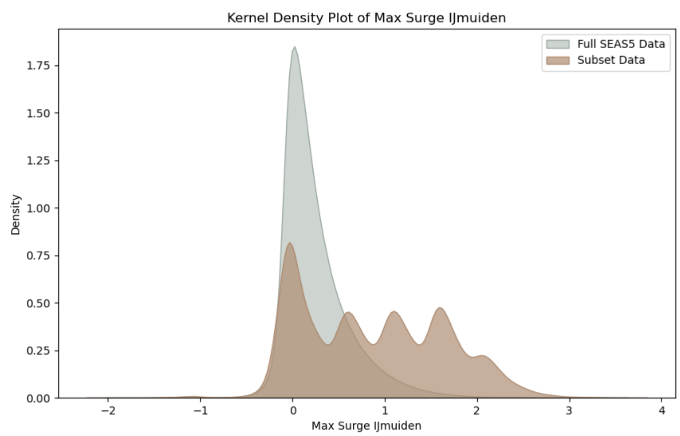


Figure 8-3: KDE plot comparing the full SEAS5 data with the subset selected using stratified sampling (61,132 days)

The key feature of this disproportionate stratified sampling method is that the same number of days is selected from each category, regardless of the number of days in each surge category. Specifically, 100,000 days were to be selected from the full dataset, and the same number of days was to be sampled from each category. For bins with fewer days than the number of samples required, all available days were selected. This resulted in a total of 61,132 selected days from the full dataset, offering a much larger sample size than MDA while maintaining a balanced representation of the surge conditions. Figure 8-3 illustrates the resulting distribution of the subset data. The stratified sampling approach was far more computationally efficient than MDA, as it could be performed on the full SEAS5 dataset without the computational burden of iterative dissimilarity calculations.

8.4 Comparison of methods

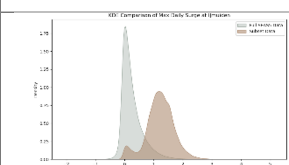
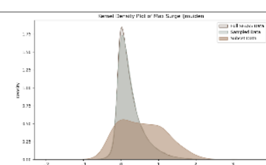
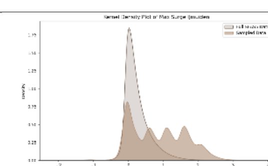
Method A results in a dataset that is skewed towards high-surge events, with fewer moderate and calm days. This is expected, as selection is based on extreme surge occurrences, with only a small fraction (10%) of additional calm conditions included. The resulting KDE plot shows a dominant peak in the higher surge ranges, reflecting the method’s emphasis on stormy conditions. However, this selection strategy leads to a lower representation of moderate surge levels, which could make it more difficult to distinguish between storms of different intensities. If the goal is to capture a full range of storm severities, this dataset may have limitations in resolving gradual transitions between calm and extreme conditions. The impact of this imbalance on storm classification remains uncertain, but it could lead to clusters that primarily distinguish between extreme storms and non-storm days, with less differentiation between varying storm intensities.

Method B, due to its selection process prioritising maximum dissimilarity, produces a more dispersed dataset. The KDE plot confirms this, showing a relatively flat distribution with no dominant peaks, suggesting that the dataset captures a broad range of surge conditions rather than emphasising any particular category. However, a major concern is that by focusing on selecting maximally different days, too many extreme surge events may be lost. Since the goal is to improve mid-term forecasting of extreme storms, this raises the question of whether MDA removes the very cases that are most relevant for storm classification. In addition to this, the small sample size of 1,000 days presents another potential issue. This sample size seems insufficient for capturing the full range of storm characteristics. With such a limited selection, there is a risk that important storm patterns and transitional weather conditions are underrepresented. With a larger sample, MDA could potentially be more useful, but in its current form, it appears less suitable for this application.

Method C results in a more balanced dataset, ensuring that all surge ranges are equally represented. Unlike Method A which strongly emphasises high surges, this approach distributes selection across different surge intensities, leading to a dataset that includes both extreme and moderate conditions. The KDE plot shows a higher proportion of very high surges (>2m) compared to the other two methods. This is because, for the highest surge strata, the total number of available days was lower than the target sample size for those categories, so all available high-surge days were selected by default. As a result, this method inherently retains all extreme surge events, ensuring that they are fully represented in the dataset. Additionally, the increased presence of lower surge conditions could help differentiate between different storm intensities, potentially improving the resolution of storm classification.

Table 8-1 summarises the different parameters of the three data selection approaches.

Table 8-1: Comparison of the three data selection approaches.

Method	Surge threshold (A)	MDA (B)	Stratified Sampling (C)
Data selection	Surge > 1.5 [m] + 10% calm conditions.	Surge at IJmuiden; MDA applied.	Surge at IJmuiden; stratified surge categories.
Selected days	123,680	1000	61,132
% of full SEAS5 data	3.6%	0.03%	1.8%
PCA components	185	130	165
Number of clusters, *k*	40	20	40
KDE plot			

9

Evaluation of generated weather patterns

This chapter presents the results of the self-clustered weather patterns and their associated storm surge characteristics. The goal is to evaluate the effectiveness of the clustering methods used to define representative weather patterns for the Dutch. Three methods were tested: surge thresholding (Method A), Maximum Dissimilarity Algorithm (Method B), and stratified sampling (Method C).

First, the generated weather patterns are visualised, along with an assessment of their distribution across different clusters. The relationship between these patterns and surge levels is explored through boxplots and heatmaps, providing insights into how different atmospheric conditions influence surge behaviour.

To further assess the quality of the weather patterns, an entropy analysis is performed to quantify the variability of storm surge values within each pattern. Additionally, Kullback-Leibler (KL) divergence is used to compare the probability distributions of storm surge values across patterns, identifying similarities and redundancies.

9.1 Weather patterns from Method A

This section presents the weather patterns generated using Method A – the surge threshold method. This approach selects days with significant storm surge levels to identify patterns associated with extreme conditions. For details on the data selection criteria used for this method, refer to Section 8.1.

9.1.1 Generated weather patterns

Figure 9-2 presents the Mean Sea Level Pressure (MSLP) patterns (left) and the corresponding gradient patterns (right) obtained from Method A. The MSLP patterns illustrate the large-scale pressure systems associated with different weather patterns, where red shading indicates high-pressure areas and blue shading represents low-pressure systems. The gradient patterns highlight pressure differences, which drive wind patterns and influence storm surge dynamics. For a more detailed view of these weather patterns, refer to Appendix D: Generated weather patterns, which provides larger versions of these figures for improved visualisation.

The distribution of cluster assignments is shown in Figure 9-1. This histogram displays the number of days assigned to each weather pattern, distinguishing between the full dataset and the subset of calm condition days included in the clustering. The distribution reveals that certain patterns occur more frequently, while others represent less common atmospheric states.

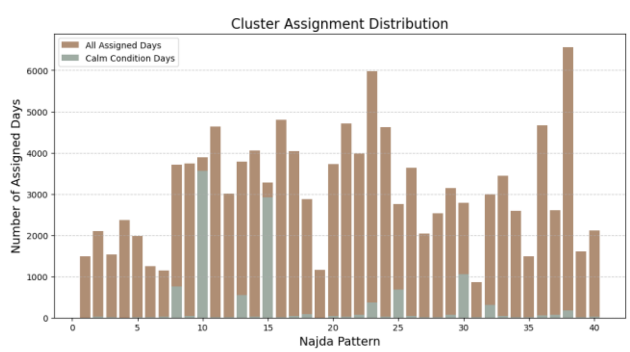


Figure 9-1: Distribution of cluster assignments across the 40 weather patterns. The grey bars indicate calm condition days included in the dataset.

Notably, the calm condition days are concentrated within a limited number of clusters rather than being evenly spread across all patterns. This is an expected and desirable outcome, as it suggests that these clusters represent predominantly calm atmospheric conditions. If the calm days were widely distributed among all clusters, it would reduce the contrast between extreme and non-extreme patterns, making it more difficult to accurately link weather patterns to their associated storm surge distributions.

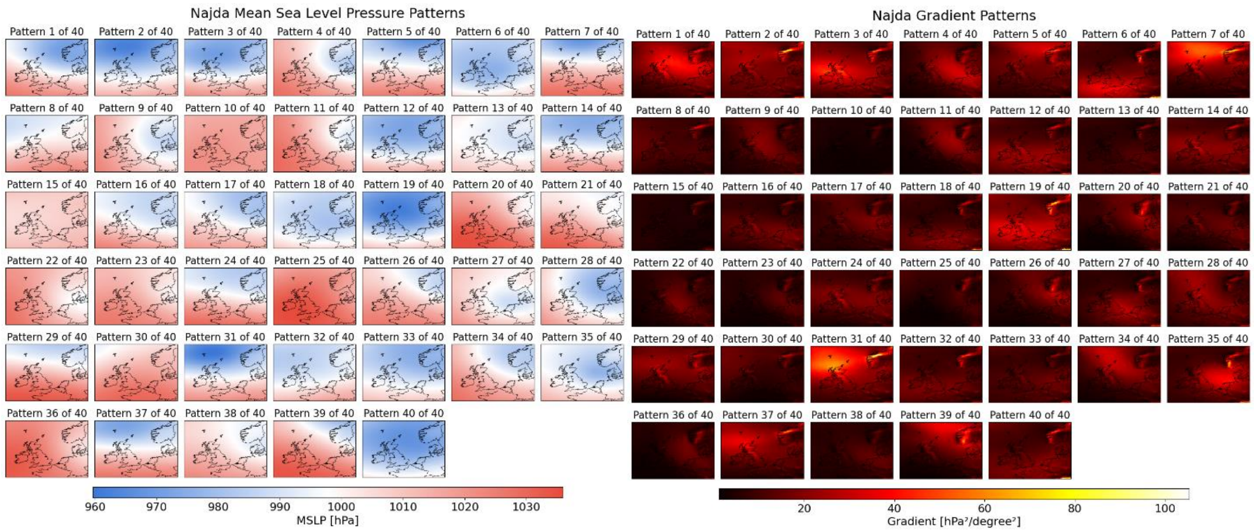


Figure 9-2: Mean sea level pressure (left) and pressure gradient (right) for the 40 weather patterns generated using Method A.

9.1.2 Surge distribution per weather pattern

The relationship between the newly generated weather patterns and storm surge values is examined by analysing the distribution of daily maximum surge values across the 40 Najda weather patterns (Figure 9-3). At shorter lead times, such as 1 day, median surge values vary across patterns, with patterns 10 and 15 exhibiting significantly lower medians, as expected given their association with calm conditions. This variability suggests that certain atmospheric patterns are more strongly linked to high surges. Compared to the Neal weather patterns assigned to the Dutch coast (Figure 6-5), where median surge values remain within a narrower range (0–0.5 m), the Najda patterns show greater distinction, indicating a clustering approach that better captures surge-relevant differences in atmospheric conditions.

Despite this improvement, most Najda patterns still exhibit a wide range of surge values, as reflected in the extended whiskers of the box plots. This large spread suggests that even within patterns typically associated with low surges, occasional high surges may still occur, posing challenges for operational applications. As lead time increases, distinctions between patterns diminish. By 15 days, median surge values converge, and interquartile ranges broaden, illustrating reduced predictive skill at longer forecast horizons. This pattern aligns with previous findings for the Neal patterns, where the association between atmospheric conditions and surge weakens over time.

To further illustrate the surge probabilities associated with each Najda pattern, Figure 9-4 presents an alternative visualisation of the surge distributions. Here, the probability of surge values falling within predefined categories for lead times of 1 and 15 days are displayed. This type of visualisation may be useful for operational forecasting, as it provides a probabilistic estimate of surge levels associated with each weather pattern. For lead time = 1 day, some patterns exhibit a high probability of surges exceeding 1 meter, whereas others, such as patterns 10 and 15, have nearly all surge values confined to the lowest category (0-50 cm). However, for lead time = 15 days, nearly all patterns exhibit a similar probability distribution, reinforcing the conclusion that these weather patterns lose their predictive skill for storm surge events at longer forecast horizons.

An important consideration is that Method A was specifically designed to select high-surge days, yet this is not strongly reflected in the probability distributions shown in Figure 9-4. While certain patterns display an increased likelihood of higher surges, the overall distributions indicate that the generated weather patterns do not focus sharply enough on high-surge events. This suggests that the selection method may not have been restrictive enough in isolating extreme surge conditions, resulting in weather patterns that still contain a significant range of moderate and low-surge days.

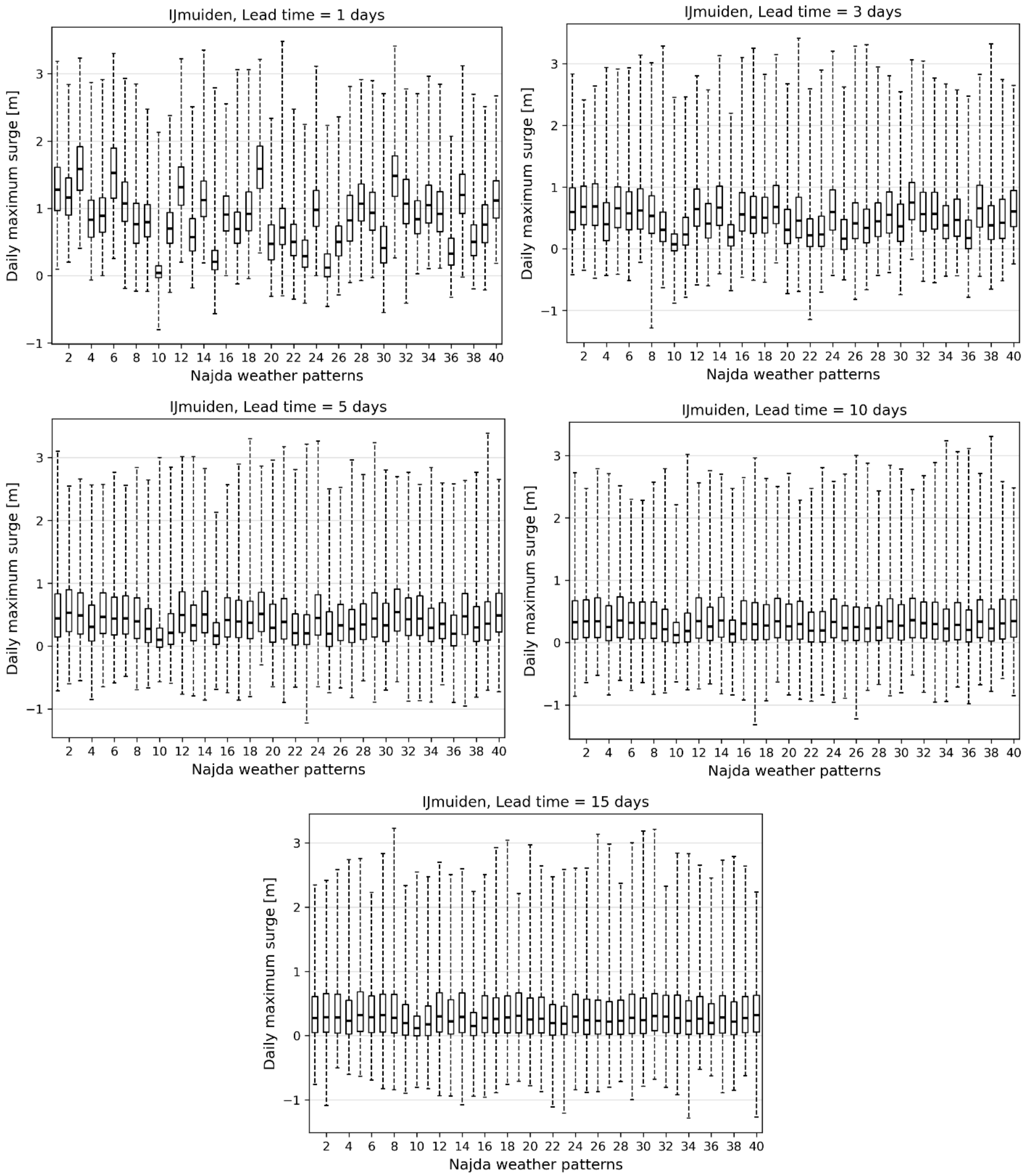
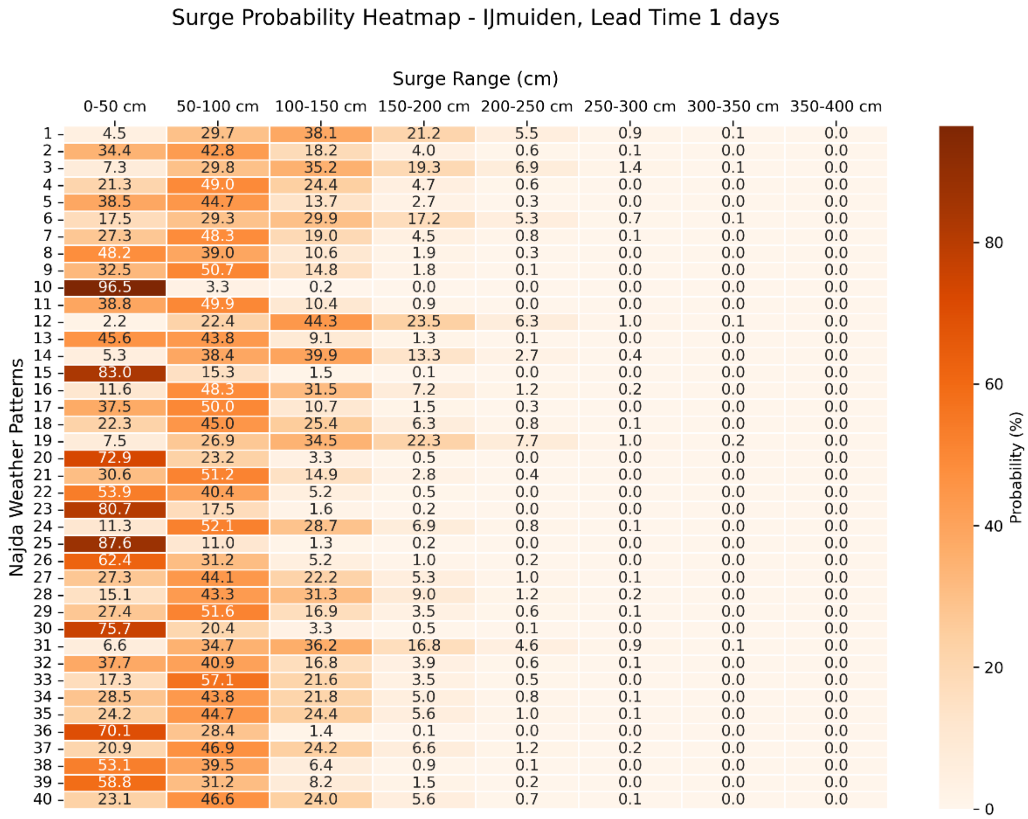


Figure 9-3: Box plots showing the distribution of daily maximum surges at IJmuiden across the 40 Najda weather patterns for lead times of 1, 3, 5, 10, and 15 days (Method A). Each box represents the interquartile range (25th to 75th percentiles), with the median shown as a horizontal line, while whiskers indicate the minimum and maximum surge values.



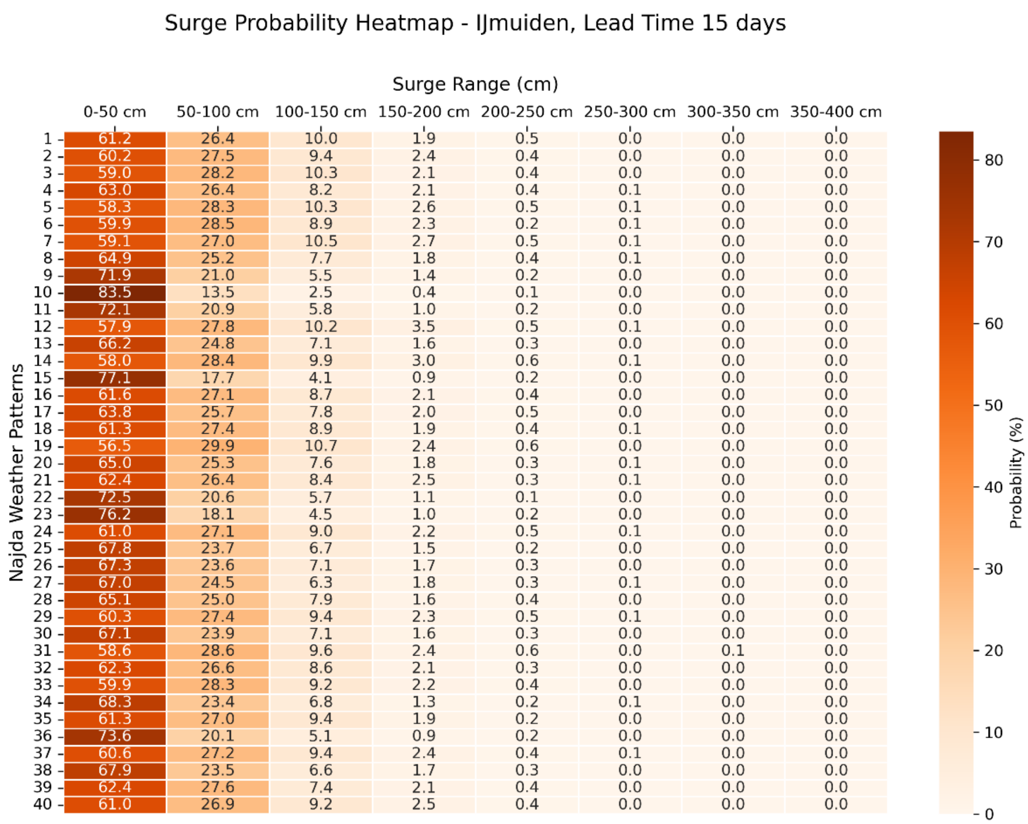


Figure 9-4: Heatmap displaying the probability distribution of daily maximum surge values at IJmuiden for lead times of 1 day (top) and 15 days (bottom).

9.1.3 Shannon entropy analysis

To assess the variability of storm surge distributions within each Najda weather pattern, an entropy-based analysis is performed. Entropy quantifies the spread of surge values, with higher entropy indicating greater variability and lower entropy suggesting more concentrated distributions. A detailed explanation of the entropy computation methodology is provided in Appendix B: Theory explained (section B.7)

Figure 9-5 presents the Shannon entropy values for each Najda weather pattern at IJmuiden for a 1-day lead time. The results illustrate that some patterns exhibit significantly higher entropy, indicating substantial variability in storm surge values, whereas others have lower entropy, signifying more consistent surge behaviour within those patterns. A more comprehensive overview, including entropy values for all clustering methods (Methods A, B, and C) and for a 15-day lead time, is provided in Appendix H: Entropy analysis.

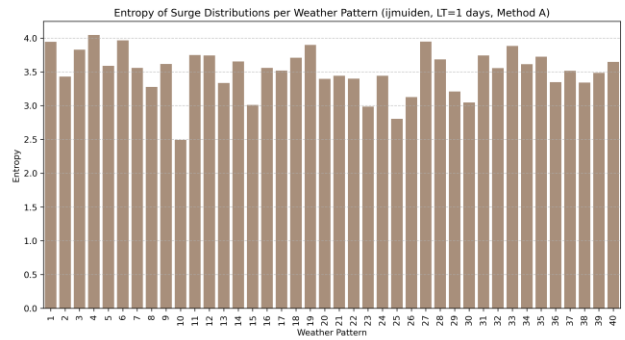


Figure 9-5: Entropy of surge distributions per Najda weather pattern for Method A. Higher entropy indicates greater surge variability.

To illustrate the contrast between high and low entropy patterns, Figure 9-6 compares the surge distributions for two representative patterns. The left panel displays the probability density functions (PDFs), where Pattern 4 (high entropy) shows a broad and dispersed distribution, indicating a diverse range of surge values. In contrast, Pattern 10 (low entropy) exhibits a sharply peaked distribution, signifying that surge values within this pattern are relatively consistent and confined to a narrow range. The right panel presents the empirical cumulative distribution functions (ECDFs), where Pattern 10 has a steep slope, indicating that most surge values are concentrated within a small range. In contrast, Pattern 4 has a more gradual slope, reflecting a wider spread of surge values.

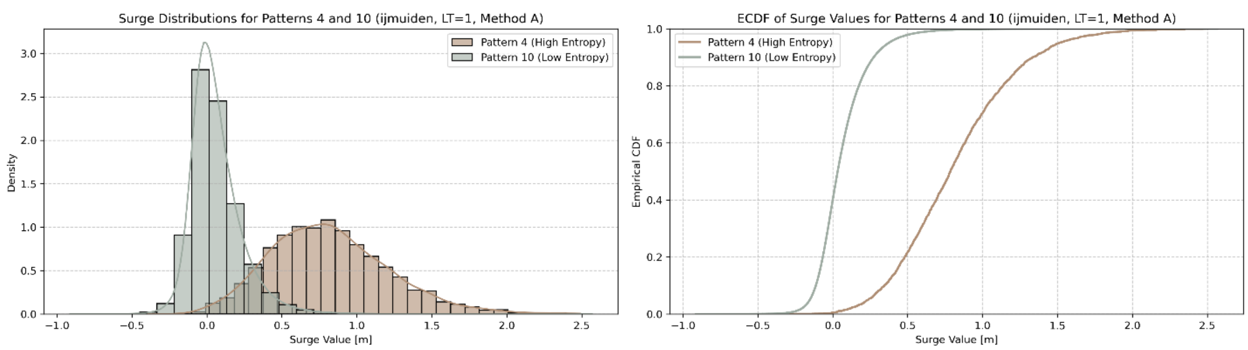


Figure 9-6: Surge distributions for Najda patterns 1 and 10 (Method A, 1-day lead time), illustrating the difference in entropy. The left panel shows the probability density functions (PDFs), where Pattern 1 has a broad, dispersed distribution (high entropy), while Pattern 10 is tightly concentrated (low entropy). The right panel presents the corresponding ECDFs.

Since the objective is to generate well-separated, distinct weather patterns, lower entropy within clusters is desirable. High-entropy patterns indicate substantial internal variability, making it difficult to associate a given pattern with a well-defined surge range. In contrast, low-entropy patterns suggest that specific atmospheric conditions lead to more predictable surge outcomes. While some variability is expected, excessive entropy within a pattern reduces its practical value for operational forecasting, as it implies that a broad range of surge magnitudes may occur under similar atmospheric conditions.

9.1.4 Relative entropy analysis

The Kullback-Leibler (KL) divergence, also known as relative entropy, quantifies the difference between probability distributions. In this context, KL divergence is used to assess how distinct the surge distributions are between different Najda weather patterns. A higher KL divergence indicates that two patterns produce significantly different surge distributions, while a lower KL divergence suggests that the patterns yield similar surge behaviour. For further theoretical details on KL divergence, refer to Appendix B: Theory explained (section B.8).

Figure 9-7 presents the KL divergence matrices for Method A at lead times of 1 and 15 days. Darker squares represent higher KL divergence values, indicating greater distinctions between the corresponding patterns. At a lead time of 1 day, several patterns exhibit strong distinctions, particularly pattern 10, which contains the majority of calm condition days. This patterns shows high KL divergence when compared to patterns associated with extreme storm surge events, reflecting its fundamentally different surge characteristics. However, as lead time increases to 15 days, the differences between patterns diminish, and the KL divergence values become uniformly lower, illustrating the decreasing predictive value of weather patterns at longer time scales.

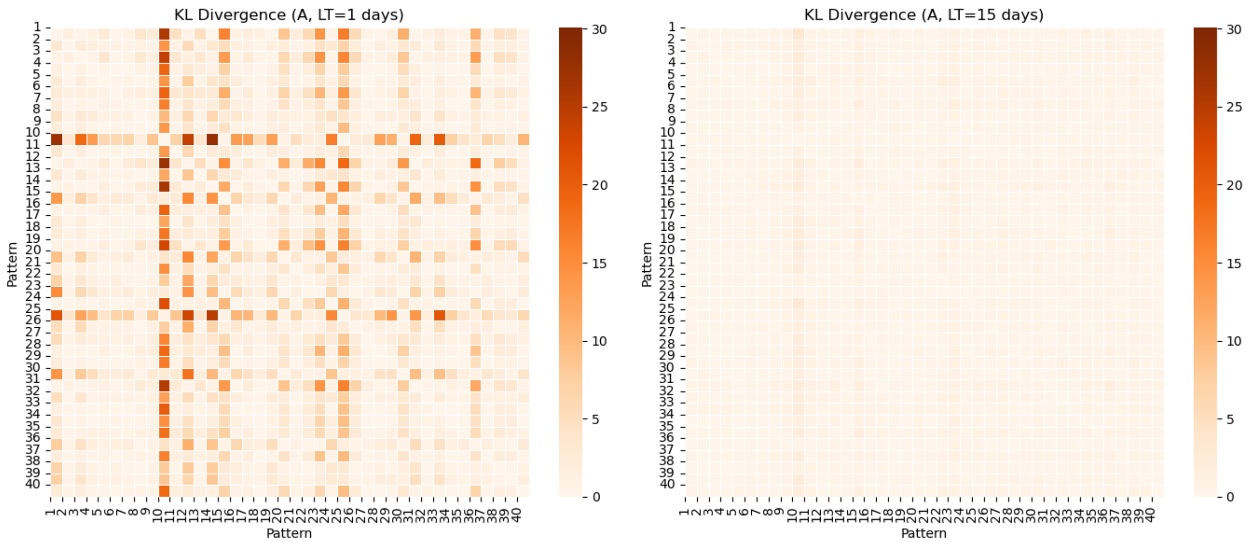


Figure 9-7: KL divergence between Najda weather patterns for Method A. Left: 1-day lead time, Right: 15-day lead time. High values (dark squares) indicate distinct surge distributions.

Figure 9-8 examines two patterns that exhibit the lowest KL divergence. The probability density functions (left panel) and empirical cumulative distribution functions (right panel) reveal that these patterns yield nearly identical surge distributions, despite originating from distinct atmospheric conditions (Figure 9-2).

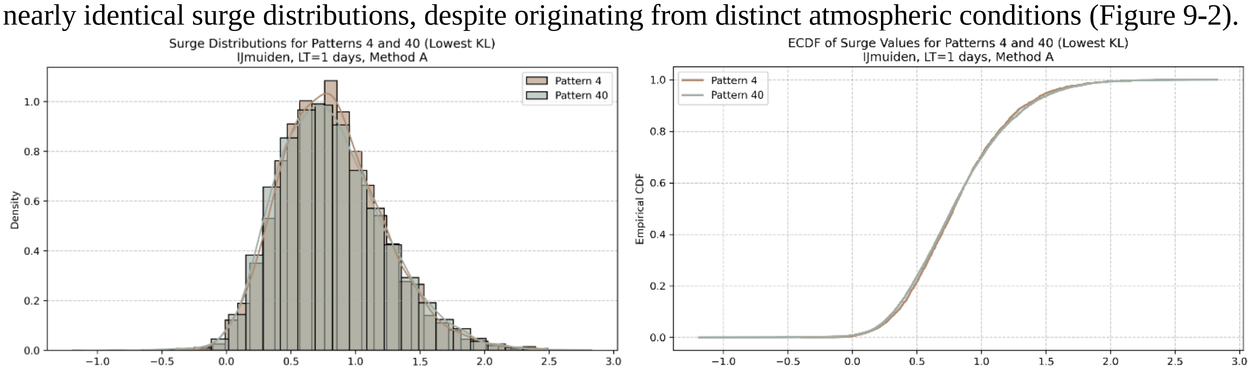


Figure 9-8: Surge distributions for Najda patterns 4 and 40 (Method A, 1-day lead time), illustrating the lowest KL divergence. The left panel shows the probability density functions (PDFs), while the right panel presents the corresponding ECDFs, both indicating highly similar surge distributions.

Figure 9-9 presents a dendrogram illustrating the clustering structure for Method A at a 1-day lead time. The vertical axis represents cluster distance, indicating the level of dissimilarity between patterns before they are merged into larger groups. For full-scale dendrograms and a more detailed explanation of hierarchical clustering, refer to Appendix E: Dendrograms.

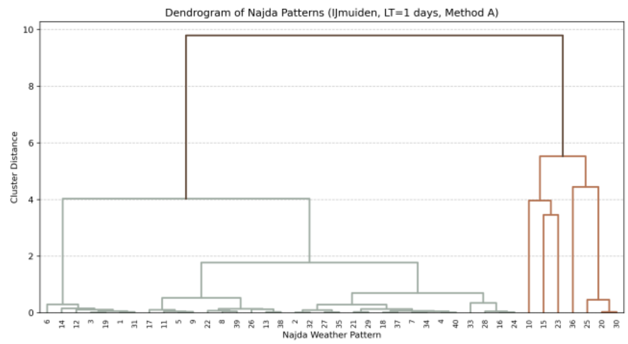


Figure 9-9: Hierarchical clustering dendrogram for Najda patterns (Method A, 1-day lead time).

A clear separation between two primary groups is visible. The brown cluster corresponds to patterns predominantly associated with calm conditions, including patterns 10 and 15, which were previously identified as having the lowest median surge values. The grey cluster contains the remaining patterns, which exhibit more extreme surge conditions. This separation reinforces the earlier findings from the KL divergence analysis, where patterns associated with low surge values showed strong distinctions from those linked to higher surges.

Conclusions

This method was specifically designed to focus on storm surge conditions while still incorporating a subset of calm days to prevent misclassification when applied to the full SEAS5 dataset. The results indicate that this objective has been partially achieved.

While certain patterns, particularly those associated with calm conditions, are clearly distinguishable, many patterns exhibit overlapping surge distributions, leading to redundancy within the classification. This lack of differentiation suggests that the clustering method used in Method A may not be fully optimised to separate surge-relevant atmospheric states.

The presence of clearly distinguishable calm-condition patterns, as seen in both the KL divergence analysis (Figure 9-7) and hierarchical clustering (Figure 9-9), suggests that the methodology effectively separates low-surge conditions from higher-surge events. Patterns 10 and 15, which contain the majority of calm days, exhibit strong KL divergence when compared to patterns associated with extreme surge events. This confirms that the inclusion of calm days in the clustering process was successful in ensuring that lower surge conditions are properly accounted for.

However, while distinct calm-condition patterns emerge, the separation among extreme surge patterns appears less pronounced. Many patterns exhibit similar surge distributions, with overlapping interquartile ranges (Figure 9-3) and low KL divergence values (Figure 9-8). This suggests that while the clustering approach captures a spectrum of surge conditions, it may not be fully optimised to differentiate between different types of high-surge events. Ideally, the patterns should reflect meaningful differences in atmospheric conditions that lead to varying surge responses, yet some patterns with distinct meteorological structures yield nearly identical surge distributions.

9.2 Weather patterns from Method B

This section presents the weather patterns generated using Method B – the Maximum Dissimilarity Algorithm (MDA). This method selects days by maximising their dissimilarity in terms of storm surge values, ensuring that both extreme and unique surge events are well-represented. For details on the data selection process and rationale behind this approach, refer to Section 8.2.

9.2.1 Generated weather patterns

Figure 9-10 presents the Mean Sea Level Pressure (MSLP) patterns (left) and their corresponding pressure gradient fields (right) obtained using Method B. Unlike Method A, which utilised 40 patterns, this approach employs only 20 patterns, as this was determined to be the optimal number of clusters based on the Within-Cluster Sum of Squares (WCSS) and Between-Cluster Sum of Squares (BCSS) criteria (see Section 7.5).

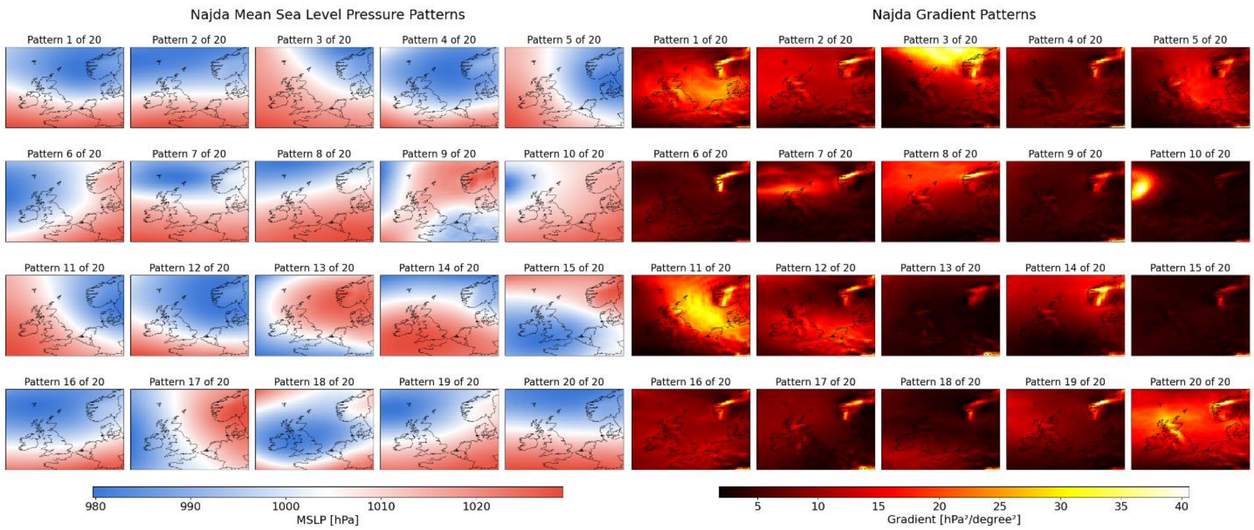


Figure 9-10: Mean sea level pressure (left) and pressure gradient (right) for the 20 weather patterns generated using Method B.

Figure 9-11 presents the distribution of cluster assignments across the 20 weather patterns generated using Method B. Unlike Method A, which incorporated a broader dataset with both high and low surge conditions, the Maximum Dissimilarity Algorithm (MDA) was applied to a much smaller subset of the SEAS5 dataset due to computational constraints.

As a result, the distribution of assigned days across patterns is notably uneven. Some patterns, such as Patterns 7 and 10, contain very few assigned days, while others, like Patterns 1 and 9, have significantly more. This imbalance suggests that some patterns represent more commonly occurring atmospheric configurations, while others capture rarer, more extreme conditions. The presence of sparsely populated clusters highlights a limitation of working with a reduced dataset: although MDA ensures a diverse range of surge conditions, the constrained selection process results in certain patterns being underrepresented.

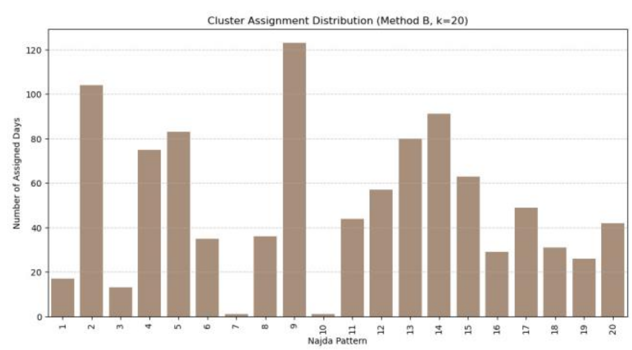


Figure 9-11: Distribution of cluster assignments across the 20 weather patterns for Method B.

9.2.2 Surge distribution per weather pattern

The distribution of daily maximum surge values across the 20 Najda weather patterns generated using Method B is shown in Figure 9-12. At a lead time of 1 day, there is noticeable variation in median surge values across patterns, though most exhibit a wide range of surge values. Unlike Method A, some patterns in Method B extend to significantly lower surge values, reaching as low as -2 meters. This is a direct consequence of the Maximum Dissimilarity Algorithm (MDA), which selects days to maximise variability, ensuring that both extreme high and low surge events are represented.

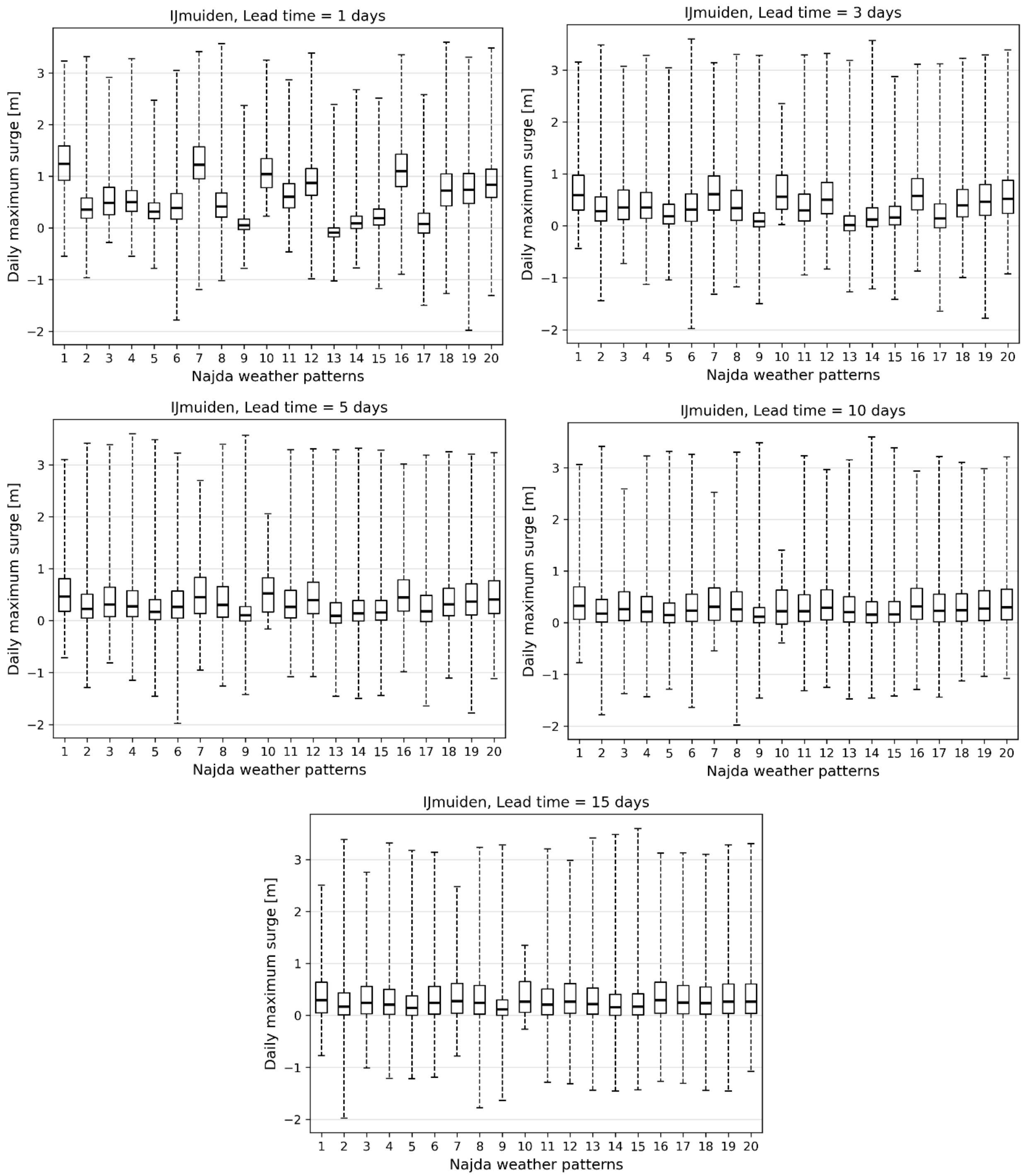


Figure 9-12: Box plots showing the distribution of daily maximum surges at IJmuiden across the 20 Najda weather patterns for lead times of 1, 3, 5, 10, and 15 days (Method B). Each box represents the interquartile range (25th to 75th percentiles), with the median shown as a horizontal line, while whiskers indicate the minimum and maximum surge values.

9.2.3 Shannon entropy analysis

The Shannon entropy values for surge distributions across the 20 Najda weather patterns generated using Method B are presented in Figure 9-13. The entropy values are consistently high across all patterns, indicating that within-pattern variability remains significant. This suggests that while the MDA ensures diverse surge conditions across all patterns, it does not necessarily optimise intra-cluster homogeneity. In other words, the method captures a broad range of surge behaviours, but individual patterns may still contain a mix of different surge conditions, limiting their ability to produce well-separated, predictable surge distributions.

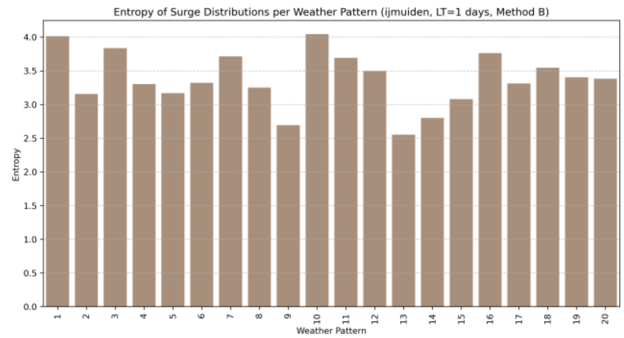


Figure 9-13: Entropy of surge distributions per Najda weather pattern for Method B.

9.2.4 Relative entropy analysis

The Kullback-Leibler (KL) divergence matrix for Method B is shown in Figure 9-14, illustrating how distinct the surge distributions are between different Najda weather patterns. At a lead time of 1 day, several patterns exhibit high KL divergence (darker squares), indicating strong differences in their surge distributions. Compared to Method A, the divergence values appear more pronounced for some patterns, reflecting the impact of the Maximum Dissimilarity Algorithm (MDA), which intentionally selected a more varied set of days.

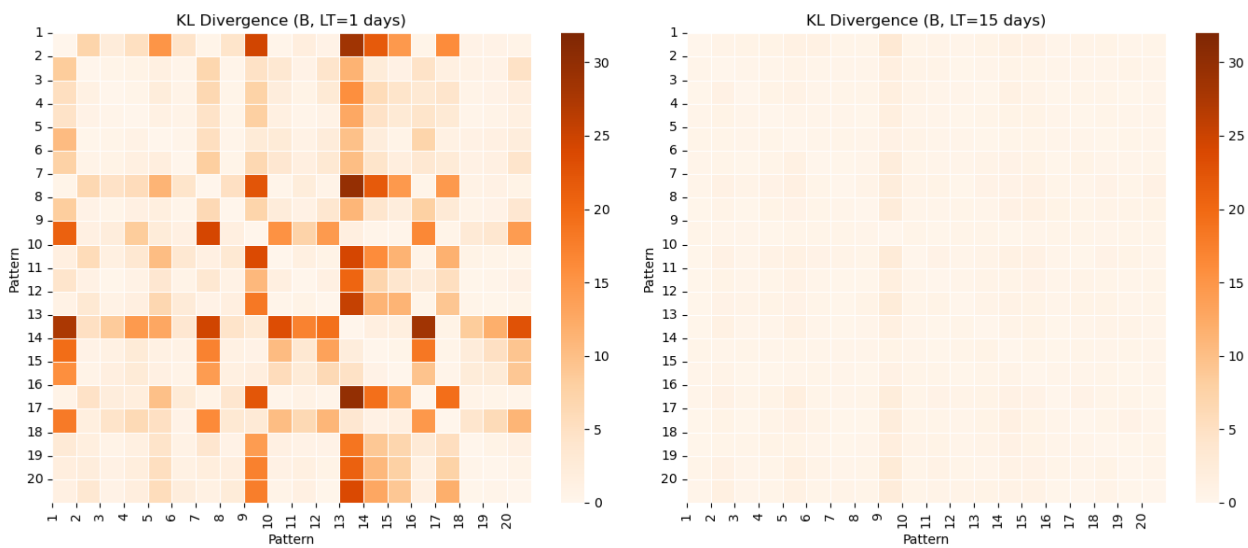


Figure 9-14: KL divergence between Najda weather patterns for Method B. Left: 1-day lead time, Right: 15-day lead time. Divergence decreases with longer lead times.

Figure 9-15 presents the hierarchical clustering dendrogram for Method B. Unlike Method A, where two primary groups (calm and extreme conditions) were clearly distinguishable, Method B exhibits a more complex clustering structure. While some patterns are still grouped closely together, suggesting similar surge distributions, the overall structure is less dichotomous. This is likely a result of the MDA selection process, which maximised dissimilarity rather than explicitly separating calm and extreme surge days.

The findings suggest that Method B produces patterns that are more evenly spread in terms of surge characteristics, as reflected by the wider distribution of KL divergence values. However, the overlap between some patterns indicates that certain atmospheric states still yield similar surge distributions.

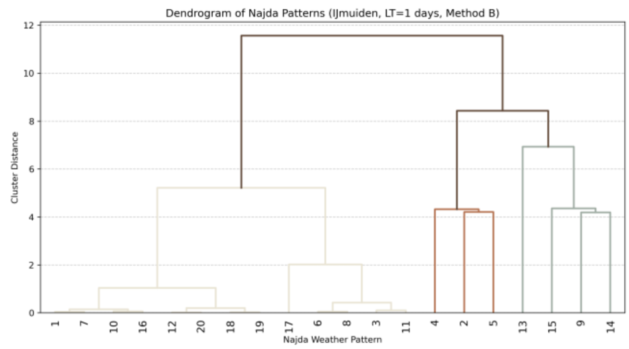


Figure 9-15: Hierarchical clustering dendrogram for Najda patterns (Method B, 1-day lead time).

Conclusions

Method B was designed to maximise diversity in surge conditions through MDA. The results confirm that this approach effectively captures a broad range of surge behaviours, ensuring that extreme and unique events are well-represented.

However, the high Shannon entropy values observed across all patterns (Figure 9-13) suggest that intra-cluster homogeneity remains limited, reducing the ability to form well-separated and predictable surge distributions. The KL divergence analysis (Figure 9-14) indicates that some patterns exhibit strong distinctions, but many still overlap, highlighting challenges in differentiating similar surge conditions. In contrast to Method A, the hierarchical clustering (Figure 9-15) results show a more structured separation of surge regimes, suggesting that MDA provides a better distinction between weather patterns despite the persistence of internal variability.

9.3 Weather patterns from Method C

This section presents the weather patterns generated using Method C – the Stratified sampling approach. This method systematically selects days by dividing the dataset into surge-based categories and ensuring balanced representation across different surge levels. For further details on the data selection process, refer to Section 8.3.

9.3.1 Generated weather patterns

Figure 9-16 presents the Mean Sea Level Pressure (MSLP) patterns (left) and their corresponding pressure gradient fields (right) obtained using Method C. This approach utilises 40 weather patterns, similar to Method A, but differs in how the days are selected. Method C employs a stratified sampling technique, ensuring that the dataset includes a balanced representation of different surge levels. This method aims to mitigate biases that might arise from the natural imbalance in surge event frequencies, ensuring that both extreme and more typical weather conditions are well-represented.

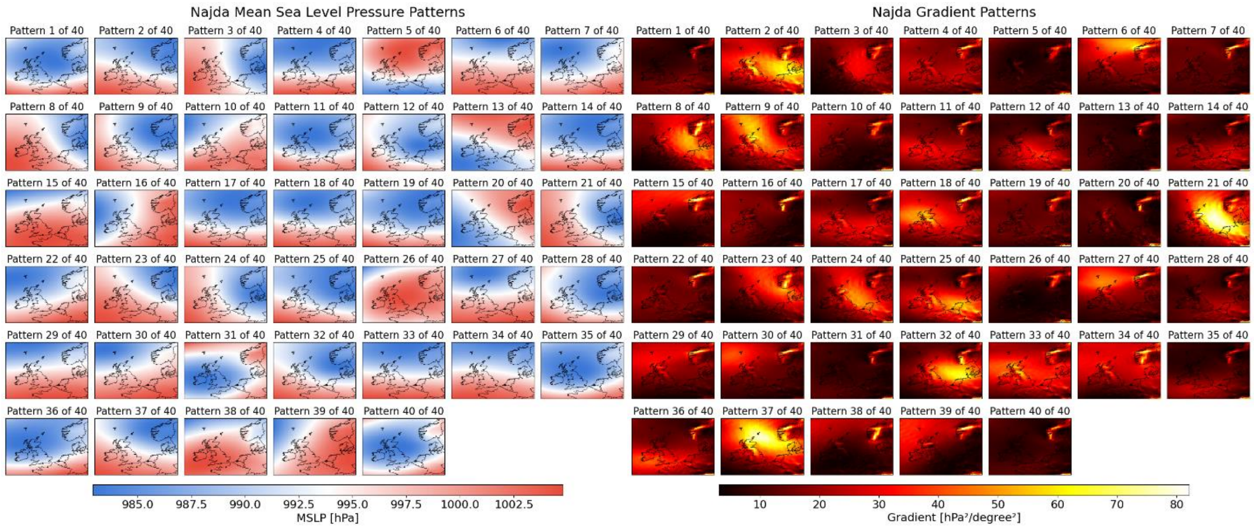


Figure 9-16: Mean sea level pressure (left) and pressure gradient (right) for the 40 weather patterns generated using Method C.

Figure 9-17 displays the distribution of cluster assignments across the 40 weather patterns generated using Method C. Several patterns, such as Patterns 4, 9, and 37, have a notably higher number of assigned days, indicating that they capture atmospheric conditions that occur frequently within the dataset. Conversely, Patterns 10, 19 and 26 contain significantly fewer assigned days, suggesting that they represent rarer atmospheric conditions or more extreme surge events.

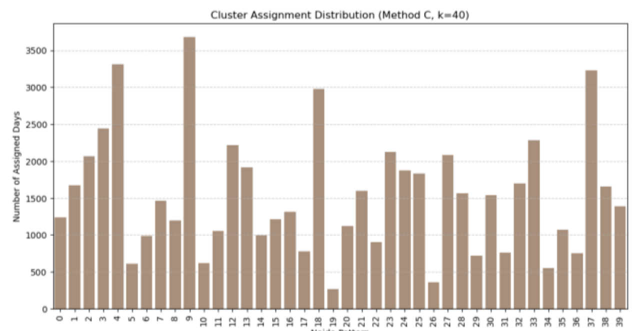


Figure 9-17: Distribution of cluster assignments for Method C.

9.3.2 Surge distribution per weather pattern

Figure 9-18 presents the distribution of daily maximum surge values across the 40 Najda weather patterns for Method C. At a lead time of 1 day, there is noticeable variation in the interquartile range (IQR) across different patterns. Some patterns exhibit narrow boxes, indicating more consistent surge values within those clusters, while others display much broader distributions, reflecting greater internal variability. This contrasts with Method A, where the interquartile ranges appear more uniform across all patterns.

As lead time increases, the distinctions between patterns diminish, similar to what was observed for Methods A and B. By 15 days, the median surge values across all patterns converge, and interquartile ranges broaden, reinforcing the finding that the predictive skill of the weather patterns weakens at longer lead times.

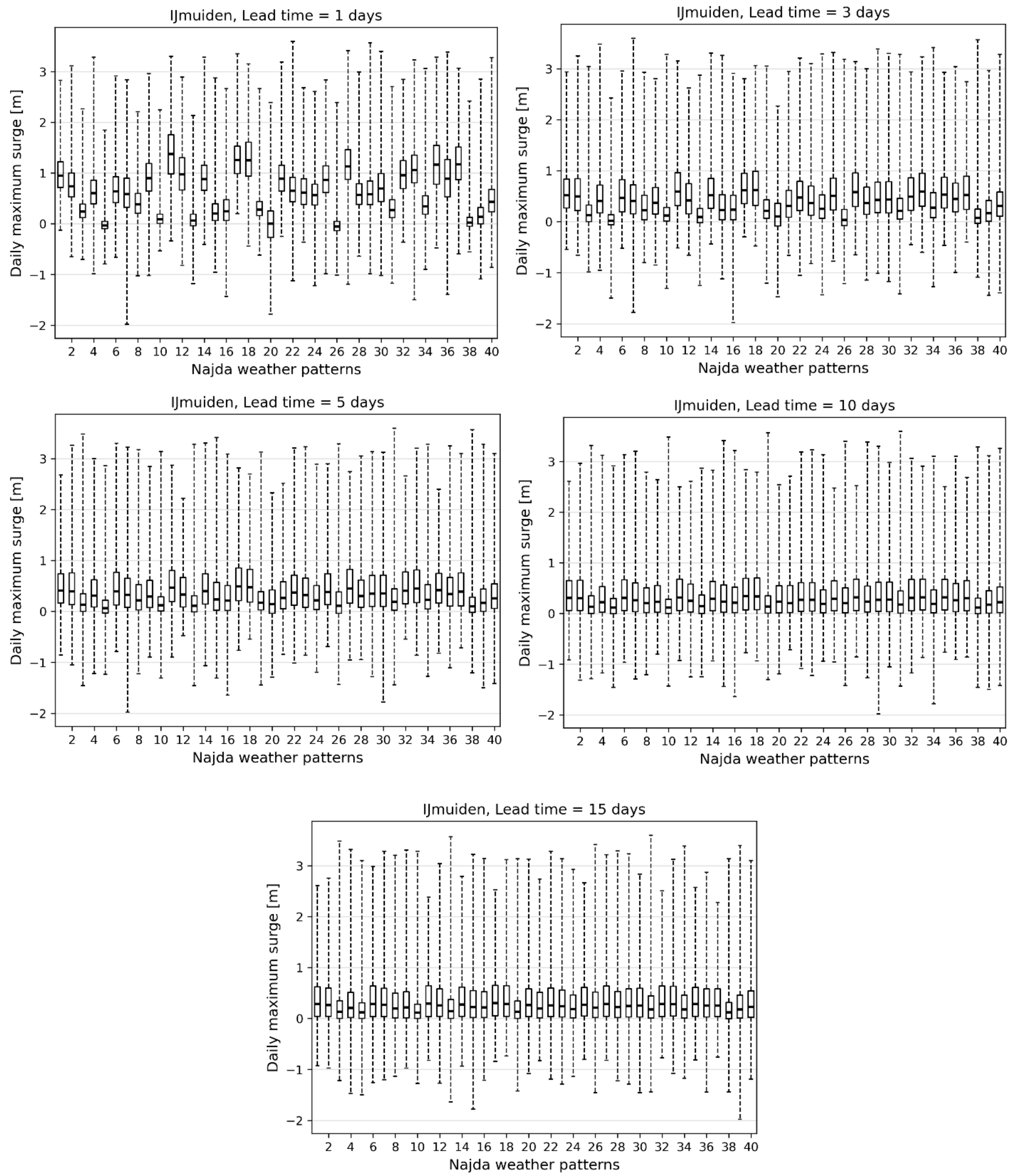


Figure 9-18: Box plots showing the distribution of daily maximum surges at IJmuiden across the 40 Najda weather patterns for lead times of 1, 3, 5, 10, and 15 days (Method C). Each box represents the interquartile range (25th to 75th percentiles), with the median shown as a horizontal line, while whiskers indicate the minimum and maximum surge values.

9.3.3 Shannon entropy analysis

Figure 9-19 presents the Shannon entropy values for the surge distributions across the 40 Najda weather patterns generated using Method C. The entropy values exhibit considerable variation, with some patterns displaying low entropy, indicating well-defined and more predictable surge distributions, while others exhibit significantly higher entropy, reflecting greater internal variability. These results indicate that while Method C improves the representation of different surge levels, it does not fully eliminate internal variability within patterns. High-entropy patterns remain, which may limit the predictability of certain weather patterns, particularly for operational forecasting applications.

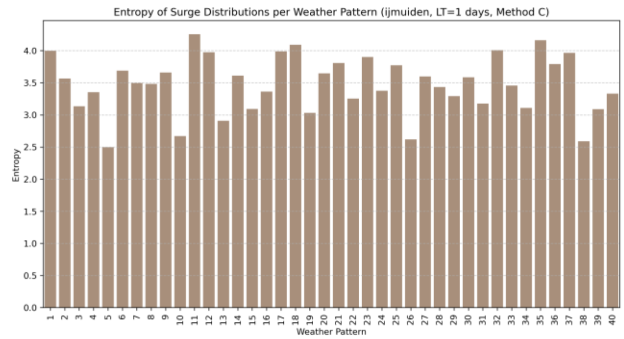


Figure 9-19: Entropy of surge distributions per Najda weather pattern for Method C.

9.3.4 Relative entropy analysis

The Kullback-Leibler (KL) divergence matrix for Method C is shown in Figure 9-20, illustrating the distinctiveness of surge distributions across the 40 Najda weather patterns. Compared to Methods A and B, more instances of high KL divergence are visible, indicating that some patterns exhibit substantially different surge distributions. However, as with the other methods, these differences fade as the lead time increases, with the KL divergence values nearly disappearing at 15 days, reinforcing the finding that these weather pattern distinctions lose predictive value at longer lead times.

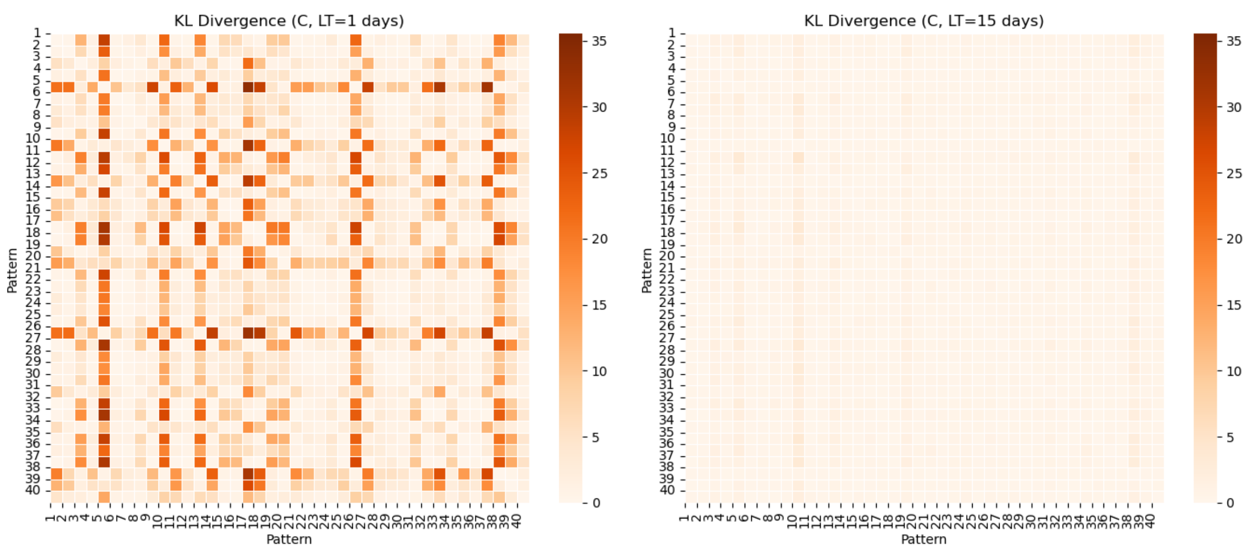


Figure 9-20: KL divergence between Najda weather patterns for Method C. Left: 1-day lead time, Right: 15-day lead time. Strong divergence suggests well-separated patterns.

To further assess the structure of the clusters generated using Method C, the hierarchical clustering dendrogram in Figure 9-21 is examined. This dendrogram reveals multiple well-defined groups, more distinct than those observed for Methods A and B. The larger number of independent branches suggests that Method C has generated patterns that are more internally coherent yet distinct from one another.

Conclusions

Method C was designed to ensure a balanced representation of different surge levels through stratified sampling. The results confirm that this approach effectively captures a diverse range of surge behaviours while maintaining more structured pattern distinctions.

The Shannon entropy analysis (Figure 9-19) shows considerable variation. Compared to Methods A and B, the KL divergence analysis (Figure 9-20) indicates stronger separation between certain patterns, suggesting improved differentiation between atmospheric states. The hierarchical clustering results (Figure 9-21) further support this, revealing clearer distinctions between groups.

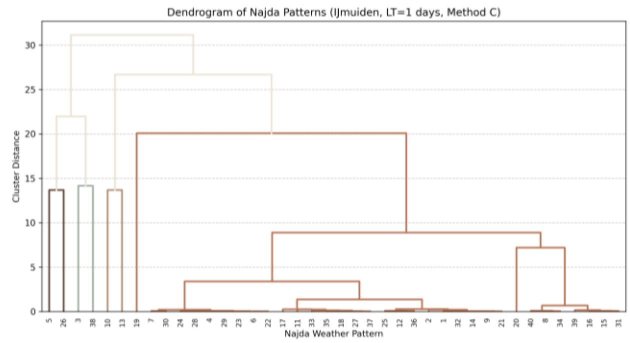


Figure 9-21: Hierarchical clustering dendrogram for Najda patterns (method B, 1-day lead time).

9.4 Observations and key takeaways

This section compares the three clustering methods (A, B and C) in terms of their ability to generate distinct and surge-relevant weather patterns. The comparison is based on the boxplots (representing the distribution of surge values across patterns), Shannon entropy (representing the internal variability of surge values within patterns) and KL divergence (representing the degree of separation between surge distributions across patterns).

The boxplots show that while all three methods capture a range of surge conditions, they differ in within-pattern variability. Method A produces relatively stable surge distributions, with clear separation between calm and extreme conditions, though some overlap remains. Method B shows more constrained surge values within individual patterns, likely due to the limited dataset used for clustering. Method C has the highest within-pattern variability, as indicated by its large IQRs, which represent a broad range of surge conditions but reduce separation between patterns.

Method A exhibits the largest differences in mean surge values, aligning with its use of extreme and calm condition days. Calm-condition patterns have narrow IQRs, while non-calm patterns show greater variability. Compared to Neal's clustering, the self-clustered methods provide clearer differentiation between patterns, though not yet distinct enough for operational use due to remaining overlap. As lead time increases, surge distributions across patterns become increasingly similar, reducing predictive value for all methods.

Method C consistently has the largest IQRs, reflecting significant internal variability. Methods A and C also capture more extreme surge events, as shown by their wider whiskers, whereas Method B, using MDA, results in fewer extreme surge cases. This suggests that B prioritises general atmospheric variability over clustering extreme events.

These findings align with the entropy analysis, where Method C shows the highest entropy values, reflecting its large within-pattern variability. Method A exhibits lower entropy in its calm condition patterns, confirming their predictability, while the surge-dominated patterns have higher variability.

KL divergence provides insight into how distinct the weather patterns are. While all methods exhibit some overlap between patterns, Method A shows strong distinctions between its calm and extreme surge patterns, making these clusters well-separated. Method C exhibits the highest KL divergence for certain patterns, meaning it successfully differentiates some weather conditions, though the high internal variability within patterns limits its overall predictive reliability. Method B results in fewer highly distinct patterns, as its selection approach focuses on maximising variability between patterns rather than creating clear-cut clusters.

Figure 9-22 provides a qualitative comparison of entropy and KL divergence across methods. The left panel illustrates the Shannon entropy values for each method at lead times of 1 and 15 days. Higher entropy values indicate greater within-pattern variability, with Method C displaying the largest spread. The right panel presents KL divergence values, showing how distinct the surge distributions are across different weather patterns. Higher KL divergence suggests better separation between patterns, though extreme values in some methods indicate isolated highly distinct patterns rather than a consistent trend across all clusters.

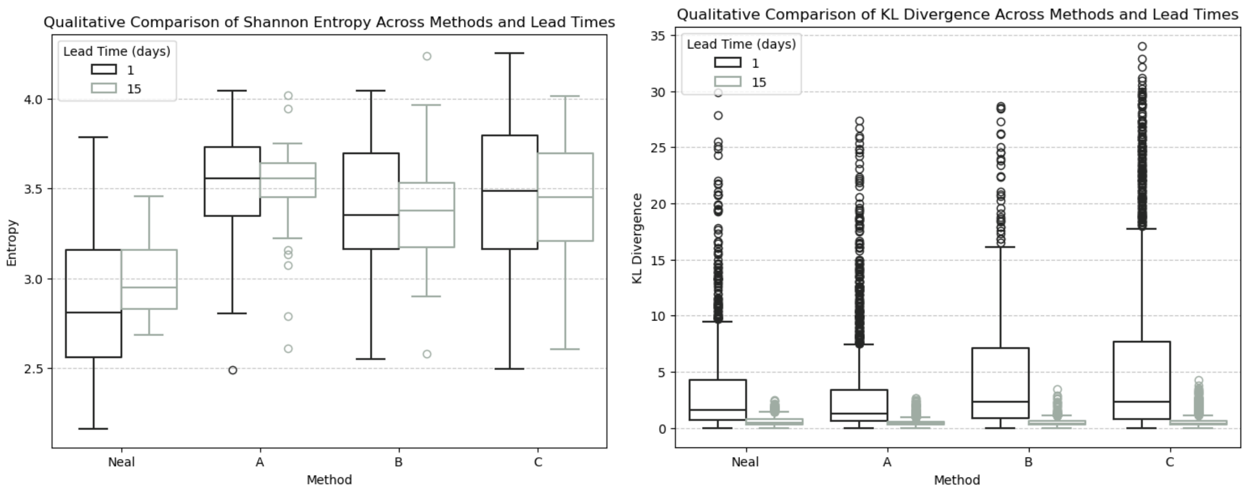


Figure 9-22: Entropy (left) and KL divergence (right) comparison across methods for both Neal and Najda weather patterns (IJmuiden, LT = 1 day and LT = 15 days).

For practical applications, these differences matter for storm surge prediction and risk assessment. Method A successfully separates calm and extreme surge conditions, making it useful when distinguishing high-impact events is a priority. Method C provides the broadest representation of surge conditions, ensuring that diverse weather events are captured, but at the cost of higher internal variability, which may reduce its usefulness for precise forecasting. Method B appears to produce patterns with lower internal variability, but it remains unclear whether this is a result of the MDA itself or the smaller dataset used for clustering. This makes it difficult to assess its suitability for practical applications without further validation. Ultimately, the choice of method depends on whether the priority is capturing a wide range of surge conditions or achieving well-separated patterns that balance internal consistency with predictive reliability.

10

Evaluation of forecasting capability: a proof of concept

The ability to accurately forecast storm surges is essential for coastal management and flood risk assessment. This chapter presents a qualitative proof-of-concept analysis investigating how well the self-clustered weather patterns predict the high surge at the peak of a representative SEAS5 storm, chosen as a proxy for storm Pia. The analysis does not aim to provide a fully quantitative validation but instead explores whether the weather patterns can offer useful early warning signals for extreme surge events.

Since storm Pia's atmospheric data was not processed in a format compatible with the weather patterns, a similar storm from the SEAS5 dataset was selected as a stand-in. The chapter first introduces storm Pia's characteristics and impact, followed by the selection process of a representative SEAS5 storm, and finally examines the predictive performance of the weather patterns (Methods A, B and C) in forecasting the peak surge of the proxy storm.

10.1 Storm Pia: characteristics and impact

Storm Pia developed in mid-December 2023 and significantly impacted Northern Europe. The storm originated near Iceland on 20 December 2023, rapidly intensifying as it moved south-eastward under the influence of a strong jet stream. Within 24 hours, its central pressure dropped from 985 hPa to 960 hPa, indicating rapid deepening. On 20 December, Pia was centred over Iceland. Less than 24 hours later, it progressed toward southern Sweden, moving quickly eastward. The storm's cold front swept across the North Sea, significantly affecting the Netherlands with strong winds and elevated water levels. The storm was particularly intense over the northern parts of the Dutch coast (Zijderveld, et al., 2024).

Pia's strong west-to-northwest winds over the North Sea led to high storm surges along the Dutch coast. Wind speeds reached 7-8 Bft (50-74 km/h) in the southern North Sea, with 9-10 Bft (75-102 km/h) over the northern North Sea and Wadden Sea. These winds, combined with high waves, caused significant water level increases at coastal locations, including Delfzijl and Harlingen. In fact, most of the main stations along the coast experienced water levels that exceeded WMCN warning levels (Zijderveld, et al., 2024).

Figure 10-1 shows the water level, tide, and surge for during storm Pia at Delfzijl, with the coloured horizontal lines indicating various warning thresholds as defined by WMCN. At Pia's peak surge in Delfzijl, the orange warning threshold was exceeded, with water level reaching 412 cm + NAP. This logs the 13th highest observed surge in Delfzijl

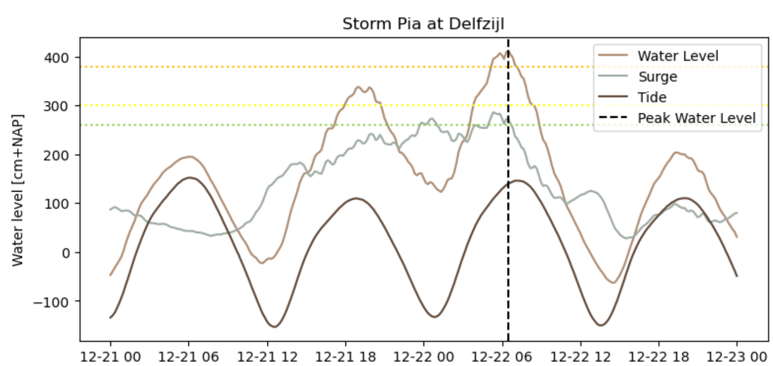


Figure 10-1: Observed water level, tide and surge at Delfzijl during storm Pia (21-22 December 2023).The green, yellow, and orange horizontal lines represent different warning thresholds defined by WMCN for water level.

since 1900. The exceedance frequency for Pia's peak water level in Delfzijl was 13 times per 100 years, reflecting the rarity of the event (Zijderveld, et al., 2024).

In response to the rising water levels, all six of Rijkswaterstaat's storm surge barriers were closed (the Maeslantkering, Oosterscheldekering, Hartelkering, Haringvlietsluizen, Hollandsche IJsselkering, and the Ramspol barrier). This event marked a significant milestone, as it was the first time that all six barriers were closed simultaneously, with the Maeslantkering being automatically operated for the first time (Rijkswaterstaat, 2023).

Following the closure of all six storm surge barriers, the impact of storm Pia on the Netherlands was largely mitigated, with no widespread catastrophic flooding occurring. However, several coastal areas did experience localised flooding. In Oudeschild on Texel, water levels of 2.7 meters above NAP inundated the harbour, and Harlingen saw water levels reach 3.2 meters, causing flooding of the quaysides. In Scheveningen, roads in the port area were briefly submerged before receding. Delfzijl, which had forecasted water levels of 4.7 meters above NAP, saw precautionary measures taken, including advisories for hotel guests to relocate (DutchNews, 2023). While the storm caused notable disruptions, the effectiveness of the storm surge barriers and early warning measures prevented more severe flooding along the Dutch coast.

10.2 Selection of a representative SEAS5 storm

Need for a representative storm

For this analysis, a SEAS5 storm was selected as a proxy for Pia due to practical constraints. The self-clustered weather patterns were designed to work with SEAS5 data, and storm Pia's raw atmospheric data was not processed into the same domain, resolution, and format as the weather patterns. Given the proof-of-concept nature of this study, using a similar SEAS5 storm was deemed sufficient to assess the predictive value of the weather patterns.

Selection process

To find the most representative SEAS5 storm, two key criteria were considered:

- The storm had to closely match Pia's water level, tide, and storm surge values at the Dutch coast.
- The storm trajectory and meteorological characteristics needed to resemble Pia's observed path and behaviour.

The initial filtering focused on hydrodynamic similarity. Pia's recorded water level peaked at 4.12 m + NAP, with a tide of 1.40 m + NAP and a storm surge of 2.72 m. Only SEAS5 storms with water level, tide, and surge values within ±2 cm of Pia's values were selected, yielding 32 candidate storms.

With this initial selection, the storm trajectory and meteorological evolution were analysed to determine how each candidate's low-pressure system, wind fields, and pressure gradients evolved over time. The goal was to identify a storm that not only matched Pia's surge characteristics but also followed a similar west-to-east track across the North Sea.

Among the 32 candidates, the synthetic SEAS5 storm on 22-23 December 1994 (from the ensemble forecast of October 1994) was identified as the best match. Table 10-1 shows that the water level, tide, and surge values of this SEAS5 storm closely resemble those of Pia at its peak. Moreover, the trajectory of this SEAS5 storm closely resembles Pia's observed movement, with rapid intensification mirrored the deepening of Pia's low-pressure system. The wind patterns associated with this storm also exhibit strong north-westerly winds over the North Sea, further reinforcing its similarity to Pia's structure. The evolution of this SEAS5 storm is illustrated in Appendix I: Trajectory comparison of storm Pia and the representative SEAS5 storm, which presents a sequence of snapshots showing its pressure contours and wind vectors at different time steps and comparing these to storm Pia.

The evolution of the selected SEAS5 storm is illustrated in Figure 10-2, which shows a comparison of the water level, tide, and surge for both storm Pia and the representative SEAS5 storm at Delfzijl. The red vertical line marks the moment of Pia's peak water level. While there are differences in the time leading up to Pia's peak, the SEAS5 storm serves as a reasonable proxy for Pia in this proof-of-concept analysis.

Table 10-1: Comparison of water level, tide, and surge at the peak of Storm Pia and the representative SEAS5 storm at Delfzijl

	Storm Pia	**Representative SEAS5 storm**
Water level	412.0 cm + NAP	411.9 cm + NAP
Tide	140.0 cm + NAP	141.4 cm + NAP
Surge	272.0 cm	270.5 cm

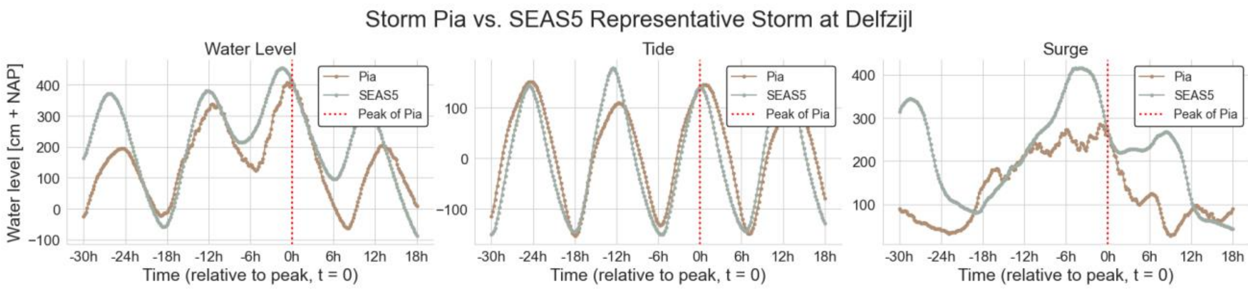


Figure 10-2: Plots of water level, tide and surge evolution of storm Pia and the representative SEAS5 storm.

10.3 Weather pattern-based prediction for the SEAS5 storm

Methodology

To assess the forecasting capability of the weather patterns, surge predictions were made for the peak surge of the representative SEAS5 storm on 1994-12-23, using the self-clustered weather patterns (Methods A, B and C). For each lead time, the surge prediction was based on the corresponding SEAS5 field from earlier dates. For example, for lead time 1, the prediction was made using the SEAS5 field from 1994-12-22, and for lead time 15, the prediction was made using the SEAS5 field from 1994-12-08. This concept may be visually viewed in Figure 5-7. Surge predictions were made up to 15 days in advance, which in this study has been defined as a mid-term forecast.

The methodology involved the following steps:

1. **Lead time matching**: The SEAS5 field for each lead time (from 1 to 15 days before the representative surge) was selected based on the date. The surge prediction was made using these SEAS5 fields. Figure 5-7 illustrates this process.
2. **Pattern matching**: The selected SEAS5 field was then matched to the closest self-clustered pattern (Methods A, B and C) via Euclidean distance.
3. **Surge prediction via boxplots**: Once the closest self-clustered pattern was identified, the surge prediction associated with that pattern was obtained from the corresponding boxplots (Figures 9-3, 9-12 and 9-18 for Methods A, B and C, respectively). The median surge and interquartile range (IQR) for each self-clustered pattern were compared to the actual surge of 2.705 m from the representative SEAS5 storm to evaluate the prediction accuracy.

Results and observations

Figure 10-3 presents a facet grid plot showing the predicted median surge for each method, with shaded IQR regions illustrating forecast uncertainty. The black dashed line indicates the actual surge of 2.705 m on 1994-12-23.

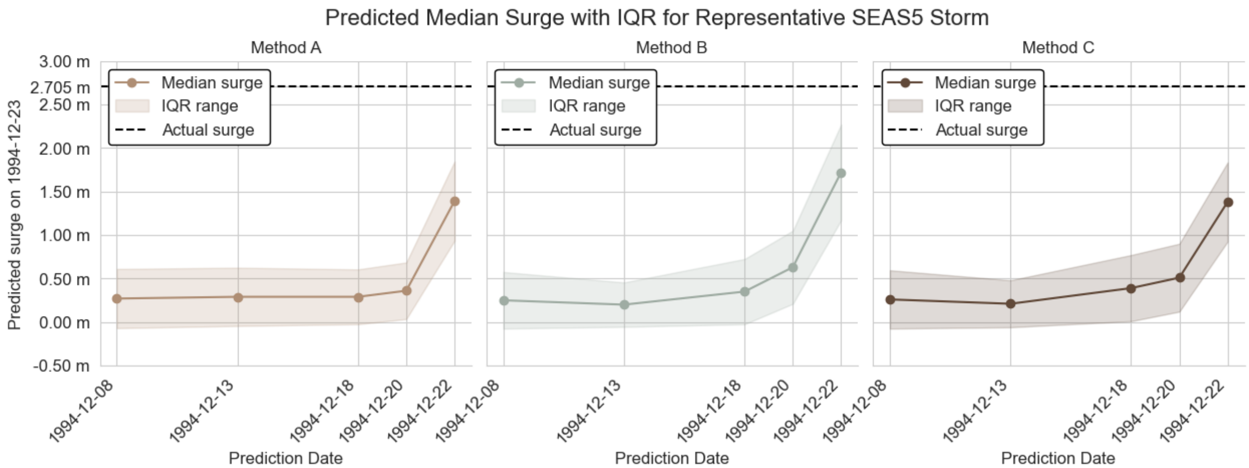


Figure 10-3: Predicted median surge with IQR for representative SEAS5 storm across Methods (A, B and C).

The results show that prediction accuracy improves as the lead time shortens. For longer lead times, the predictions significantly underestimate the surge. For example, at a 15-day lead time, the median surge predictions were as low as 0.27 m for Method A, compared to the actual surge of 2.705 m. However, even at shorter lead times (e.g., 1 day), the predictions are still notably off, with the highest predicted median surge at 1.72 m (Method B). The discrepancy between the predicted median surge and the actual surge remains large across all methods.

When comparing Methods A, B, and C, the predicted surge patterns are somewhat similar, but there are differences in the timing of when the surge predictions start to increase. Method A, for example, keeps the predicted surge low for a longer time before showing any upward trend, while Methods B and C begin to show a surge increase slightly earlier. This difference might be important when interpreting operational forecasts, where an earlier signal of a high surge could be crucial.

Uncertainty, as shown by the interquartile range (IQR), remains substantial across all lead times. The IQR fluctuates between approximately 0.5 and 1 m throughout the lead times, reflecting considerable uncertainty in the predicted surge levels.

11
Discussion

This study investigated the potential of integrating weather pattern-based classification into mid-term storm surge forecasting for the Dutch coast. The research was conducted in two phases: Phase I assessed the applicability of Neal et al. (2018) predefined weather patterns, while Phase II explored self-clustered patterns derived directly from the SEAS5 dataset.

This chapter explores several key questions: Why was this method expected to work (Section 10.1)? Why did it struggle, both for Neal's predefined patterns and the self-clustered methods (Section 10.2)? What are the practical implications for WMCN (Section 10.3)? And how does weather pattern classification fit into broader interdisciplinary applications (Section 10.4)?

11.1 Conceptual basis for weather pattern-based storm surge forecasting

The central hypothesis of this study was that large-scale weather pattern classification could enhance mid-term storm surge forecasting for the Dutch coast by identifying recurring atmospheric configurations associated with extreme surge events. This idea was motivated by the demonstrated success of weather pattern-based forecasting in the UK by Neal et al. (2018).

Reasons for applying weather pattern classification

I. ***Synoptic-scale influence on storm surges*** - The formation and intensity of storm surges along the Dutch coast are heavily influenced by large-scale atmospheric circulation patterns (Section 2.2). De Kraker (2010), has shown that low-pressure systems over the North Sea, combined with strong north-westerly winds, are primary drivers of extreme surge events in this region. By classifying these recurring synoptic configurations, the expectation was that storm surge risks could be anticipated further in advance than is currently possible with traditional numerical models.

II. ***Potential for extending forecast lead times*** - Currently, operational storm surge models provide reliable forecasts up to 10 days ahead (Section 1.3). However, extending beyond this lead time is challenging due to computational demands, a lack of flexibility in existing models and inherent atmospheric uncertainty. Weather pattern classification, if successful, could serve as a complementary tool to identify high-risk periods earlier, bridging the gap between short-term numerical forecasts and long-term climate projections.

III. ***Comparison with existing pattern-based forecasting methods*** – Neal's weather patterns, originally developed for the UK, have been successfully applied in flood risk forecasting (Section 2.5). This study aimed to assess whether a similar methodology could be adapted to the Dutch coastal context using both these predefined weather patterns from Neal et al. (2018) (Phase I, Chapter 6) and self-clustered weather pattern derived from SEAS5 data (Phase II, Chapter 9).

Despite these theoretical justifications, the results indicated that weather pattern classification did not significantly improve storm surge predictability at longer lead times. The following section analyses the key reasons for this outcome.

11.2 Assessment of predictive skill and methodological limitations

While the theoretical foundations for applying weather pattern classification to storm surge forecasting appeared promising, the results showed that neither Neal's predefined weather patterns nor the self-clustered methods were able to significantly improve predictive skill for the Dutch coast in their current form. Instead of identifying distinct atmospheric configurations that consistently preceded extreme surge events, the classifications exhibited considerable overlap between different surge magnitudes. Several factors contributed to this outcome, related to dataset choices, methodological assumptions, and fundamental characteristics of storm surge formation.

This section first evaluates Neal's predefined weather patterns and then examines the self-clustered methods (A, B, and C), analysing the methodological limitations that contributed to their weaker-than-expected performance.

11.2.1 Neal weather patterns

Several factors may have contributed to the limited predictive skill of Neal's weather patterns for Dutch storm surge forecasting, including:

I. **Spatial domain mismatch: clipping of Neal's patterns** - Neal's patterns were originally derived over a larger North Atlantic-European domain, based on how well UK temperature and precipitation time series could be reconstructed using only the weather pattern classification (Neal, Fereday, Crocker, & Comer, 2016). This domain was not explicitly optimised for storm surge forecasting, but rather for understanding UK climatological variability and weather regime transitions. In this study, the Neal patterns had to be clipped to match the smaller DCSM5 domain. As shown in Figure 5-2, this clipping may have removed crucial large-scale atmospheric structures, particularly those influencing storm development in the North Atlantic, where many extreme storms impacting the Dutch coast originate (Bell, Gray, & Jones, 2017) (Section 2.1). Thus, the domain mismatch may have weakened the ability of Neal's weather patterns to capture surge-relevant atmospheric conditions, limiting their effectiveness in the Dutch coastal context.

II. **Optimised for the UK** - Unlike this study, which aimed to use weather patterns explicitly for storm surge forecasting, Neal's method was designed as a flexible approach for post-processing ensemble forecasts in various applications, including but not limited to storm surge forecasting (Neal, Fereday, Crocker, & Comer, 2016). Additionally, the weather patterns were not explicitly tailored to Dutch storm surge characteristics, meaning that the optimal atmospheric patterns for UK flood risk may not necessarily align with those influencing Dutch surge events.

III. **Absolute MSLP vs. anomalies** - Another key difference is that Neal's clustering method used MSLP anomalies, while in this study, absolute MSLP values were applied. Anomaly-based clustering removes seasonal and large-scale biases, making it potentially better suited to identifying deviations that drive extreme weather events. By contrast, using absolute MSLP values may have resulted in weaker clustering performance, as it does not highlight deviations from climatological norms.

IV. **Differences Between EMULATE and SEAS5** - Neal's clustering was performed on 154 years of observed data (EMULATE dataset), whereas in this study, the same patterns were applied to 9000+ years of SEAS5 seasonal forecast data. While Section 6.2 showed that the mean MSLP values in EMULATE and SEAS5 were similar, the standard deviation was found to be higher in SEAS5. This suggests that SEAS5 contains a greater degree of variability, meaning that applying weather pattern classifications derived from EMULATE to SEAS5 may have introduced inconsistencies that weakened the pattern-matching process.

11.2.2 Self-clustered patterns

The goal of the self-clustering methods was to generate new weather patterns optimised for Dutch storm surge forecasting, using SEAS5-based data. However, the classifications struggled to meaningfully separate extreme surge events from moderate or low-surge cases for longer lead times, similarly to the Neal weather patterns. Several common issues affected all three self-clustering approaches, related to lead time selection, predictor choices, and spatial domain constraints.

This section first examines these general methodological challenges that impacted all self-clustered approaches. It then provides a method-specific evaluation of Methods A, B, and C, assessing their individual performance and limitations.

General analysis

The following general aspects influenced the performance of all self-clustering methods:

I. Lead time considerations: were the right days clustered?

A major limitation of the current methodology is that days were selected for clustering based on peak surge values. This means that clustering was performed on the same day as the maximum surge, rather than on the preceding atmospheric conditions that led to the event.

For storm surge forecasting, identifying precursors is more useful than classifying conditions on the surge day itself. Since the goal of pattern-based classification is to provide mid-term forecasts, the current approach may have weakened the usefulness of the resulting clusters for real-world applications.

A more effective strategy would involve:

- Selecting surge events but clustering based on the preceding atmospheric conditions, which would better capture storm evolution.
- Exploring multi-day sequences instead of clustering on individual days, as storms often take several days to develop before reaching peak surge levels.

II. Atmospheric variable selection: was MSLP sufficient?

This study relied solely on mean sea level pressure (MSLP) as the primary atmospheric variable for clustering, using two predictors: MSLP itself and its gradient. However, since the gradient is directly derived from MSLP, the clustering was effectively based on just one underlying variable. While MSLP plays an important role in defining large-scale weather patterns, other variables also influence storm surge dynamics and could improve clustering performance.

The gradient of MSLP provides indirect information about wind strength and direction, as the spacing of isobars describes wind speed, and their orientation indicates wind direction (Hegermiller, et al., 2017; Espejo, Camus, Losada, & Méndez, 2014). However, this relationship is not always direct, as Coriolis forces and surface friction modify the actual wind trajectory, particularly in the boundary layer (Holton & Hakim, 2013). Pugh (1987) highlights that in extratropical cyclones, both pressure and wind stress contribute to storm surge formation. The omission of wind stress as a predictor may therefore have limited the ability of the clustering methods to distinguish between storm patterns that generate high surges and those that do not.

III. Spatial domain constraints: was the study area too small?

As discussed in the previous section, the Neal weather patterns had to be clipped to fit the SEAS5 domain used in this study, based on the DCSM domain. Similarly, the self-clustered patterns were all derived within the same study area also.

This raises the same concern: many storms that generate extreme surges along the Dutch coast originate in the North Atlantic, which is out of scope in this study. If key storm features develop outside the clustering domain, then pattern-based classification may inherently struggle to provide useful signals at longer lead times. One way to address this would be to expand the clustering domain to include the North Atlantic, capturing storm development before it reaches the Dutch coast.

Method-specific analysis

The following sections evaluate the performance of each self-clustering method individually, assessing their strengths and limitations, and discussing how specific methodological choices influenced their ability to improve storm surge predictability.

I. Method A

Findings and insights

Method A applied a fixed storm surge threshold to classify high-surge days for clustering. This approach successfully identified a clear separation between calm conditions and storm-related patterns, as reflected in the boxplots (Figure 9-3) and dendrogram results (Figure 9-9). The calm condition patterns were well-defined, demonstrating that the method captured large-scale differences between stormy and non-stormy weather regimes.

Challenges and areas for improvement

While Method A effectively differentiated stormy from calm conditions, the use of a fixed surge threshold across all locations introduced regional inconsistencies. For instance, the average surge in Delfzijl is significantly larger than in Vlissingen (Appendix C: Surge return period analysis), meaning that the same threshold could classify a moderate event in one location but an extreme event in another.

Moreover, because all locations were included whenever the threshold was exceeded at any one station, the dataset included many surge values below 1.5 meters (beyond those added by the calm condition days), resulting in greater variability within the high-surge clusters. This contributed to less distinct extreme-surge patterns, limiting their predictive value.

Possible future refinements

- Using a fixed return period for the surge instead of a fixed threshold could improve regional consistency and ensure a balanced representation of extreme events across all locations. Instead of applying a universal surge threshold, the selection could be location-specific, ensuring that storm days are based on historical exceedance probabilities for each site.
- Performing clustering separately for each location could further improve the optimisation of weather patterns. Instead of assigning storm days to all locations when a threshold is exceeded at just one site, clustering could be conducted independently for each location, ensuring that the identified patterns are better tailored to local surge dynamics.

II. Method B

Findings and insights

Method B used the Maximum Dissimilarity Algorithm (MDA) to select a representative subset of 1000 days from the SEAS5 dataset. Due to computational constraints, MDA could not be applied directly to the entire SEAS5 dataset, which spans over 3.5 million days. Instead, the approach first randomly selected a subsample of 10,000 days from the full dataset. Then, within this randomly drawn subset, MDA was applied to further refine the selection to 1000 days, resulting in the final dataset representing a diverse range of atmospheric conditions while maintaining a mix of high-surge and low-surge events.

Interestingly, despite working with a subsample rather than the full SEAS5 dataset, the randomly selected 10,000-day subset retained a very similar surge distribution to the full dataset. This suggests that even a reduced selection can preserve the overall statistical properties of SEAS5, making random sampling a viable preprocessing step for large-scale surge classification.

Additionally, while MDA promotes diversity in the selected cases, it still ensured that extreme surge days were adequately represented, particularly at IJmuiden, the representative location used for MDA selection. Despite selecting cases based on a single reference location, the final dataset contained a broad range of surge values across other locations as well, indicating that IJmuiden-based selection still provided a reasonably representative distribution along the Dutch coast.

Challenges and areas for improvement

The selection of only 1000 days was a major constraint, as it was purely a computational necessity rather than an optimal methodological choice. Given that SEAS5 contains over 9,000 years of data, this small sample size is likely not to have been sufficient to capture the full range of distinct weather patterns relevant to Dutch storm surges.

Another issue is that the resulting cluster assignments were highly imbalanced, with some weather patterns containing very few assigned days (<10 days). The presence of sparsely populated clusters indicates another limitation of working with a reduced dataset: although MDA ensures a diverse set of surge conditions, the constrained selection process leads to underrepresentation of certain patterns.

Additionally, MDA treats all surge categories equally, meaning that moderate and extreme surge events are weighted the same in the selection process. However, for storm surge forecasting, a more nuanced approach that prioritises high-surge conditions could be more beneficial. Assigning more weight to high-surge events during MDA selection could help generate more weather patterns focused on extreme conditions, improving differentiation within the highest surge categories.

Finally, the selection process used only a single representative location, IJmuiden, for computational reasons. While the results suggest that the selected data still provided a broad and representative surge distribution across other locations (see Appendix F: KDE plots), it remains unclear whether certain localised surge dynamics were underrepresented due to this single-location selection approach.

Possible future refinements

- Increasing the dataset size would provide better statistical robustness of clustering, ensuring that all relevant atmospheric configurations are adequately represented.
- Instead of relying solely on IJmuiden, exploring multi-location MDA selection could enhance regional representativeness and ensure that the chosen days for clustering better reflect coast-wide surge dynamics.
- Refining the MDA selection process to prioritise storm surge-relevant cases could enhance its effectiveness in storm classification.

III. Method C

Findings and insights

Method C aimed to improve dataset balance by ensuring a more even representation of high-surge and low-surge cases, preventing the clustering process from being overly dominated by calm conditions. This approach successfully resulted in a broader distribution of surge events, capturing a range of surge values across different surge categories (e.g., low, medium, high).

A key advantage of this method was that all high-surge events for IJmuiden (the reference location used for this method) were included in the dataset. Since there were fewer high-surge days available, it was not possible to perform a random selection within the highest surge strata, so instead, all high-surge days were retained. This ensures that extreme surge events were always included in the clustering process, in addition to the more moderately distributed cases from other strata.

The kernel density estimation (KDE) plots (Figure 8-3) revealed that for IJmuiden, the resulting surge distribution exhibited multiple peaks. This suggests that certain surge values were overrepresented within each category, potentially biasing the clustering process.

Another notable observation was that nearby locations (Hoek van Holland and Den Helder) exhibited similar KDE structures to IJmuiden, while more distant locations (Delfzijl and Vlissingen) showed smoother distributions (Appendix F: KDE plots). This indicates that while the stratified sampling preserved variability in surge categories at locations close to IJmuiden, it may have been less representative for more distant locations where surge behaviour differs more significantly.

Challenges and areas for improvement

Although Method C improved the balance between high-surge and low-surge cases, the fact that stratified sampling was based only on IJmuiden meant that the resulting weather patterns were optimised primarily for this location. While this still produced reasonably broad surge distributions for other locations, it is unclear whether the storm conditions most relevant for Vlissingen or Delfzijl were sufficiently captured.

The KDE analysis suggests that locations closer to IJmuiden were better represented, while locations farther away had smoother, less pronounced peaks in their surge distributions. This suggests that surge variability at distant locations may not have been fully preserved, potentially limiting the ability of the weather patterns to distinguish regionally specific storm surge dynamics.

Key lessons and possible future adjustments

- Expanding stratified sampling to include all locations rather than just IJmuiden could improve regional representativeness. By applying stratification individually at each location, rather than assuming that IJmuiden-based classification generalises well to the entire Dutch coast, the resulting weather patterns could be better optimised for location-specific surge behaviour.
- Refining the way surge categories are defined to reduce overrepresentation of certain surge values within each category could lead to more uniform distributions, preventing biases in the clustering process.

11.2.3 Key takeaways

The assessment of both Neal's predefined weather patterns and the self-clustered approaches has shown several key limitations and areas for improvement in applying weather pattern classification for storm surge forecasting:

- ***Predefined weather patterns may not generalise well across regions*** **-** Neal's patterns were optimised for UK climate conditions and flood forecasting, making them less effective for Dutch storm surges. Differences in spatial domain, dataset characteristics, and clustering methodology likely reduced their applicability.
- ***Selecting clustering days based on peak surge values may not be optimal*** **-** Instead, clustering should focus on the atmospheric conditions leading up to extreme surge events rather than the surge day itself.

- ***MSLP alone may not be sufficient*** - Incorporating additional predictors such as wind stress could improve the ability of clustering to distinguish between storm intensities.
- ***Expanding the spatial domain could improve predictive skill*** - Many storms that drive extreme surges originate in the North Atlantic, a region that was outside the SEAS5 domain used in this study. Expanding the clustering domain could improve early detection of relevant storm patterns.
- ***Sample size matters -*** Increasing the number of selected days for clustering, particularly in MDA-based methods, could prevent overfitting to a small subset of cases, improving the statistical robustness of the classification.
- ***Regional surge dynamics should be considered*** - Methods that rely on a single representative location (e.g., IJmuiden) may not fully capture surge variability across the Dutch coast. Expanding the methodology to account for regional differences could improve classification accuracy.

11.3 Practical implications for WMCN and operational forecasting

The results of this study provide insights into how weather pattern classification could be applied within WMCN's operational framework. While the classification methods did not improve longer lead-time predictability, they may still have value in certain forecasting scenarios. The usefulness of these weather patterns depends largely on the specific forecasting objective.

If the goal is to predict whether a high surge will occur 15 days in advance, weather pattern classification, as currently defined, provides little practical value. The surge distributions across different weather patterns at such long lead times show too much overlap, making it difficult to distinguish between high and low surge events. However, if the focus shifts to a shorter-term forecast within an existing ensemble prediction, such as using an SEAS5 forecast for 14 days from now to assess surge risk at day 15, then the classification becomes more relevant. At shorter lead times, the 1-day lead time boxplots show clearer distinctions between weather patterns, suggesting that such an approach could help refine surge risk assessments based on ensemble forecasts.

In situations where WMCN aims to determine only whether a high surge event will occur (binary classification), Method A appears to be the most effective. The clustering method successfully distinguished between calm and extreme surge days, as reflected in both the boxplots and dendrogram results, where two distinct pattern groups emerged. This suggests that weather pattern classification could serve as a probabilistic screening tool to flag periods of increased surge risk. However, if the goal is to estimate the severity of the surge, the weather patterns appear less suitable. Many of the identified patterns exhibit wide surge distributions, meaning that within a single pattern, surge values can vary by several meters. As a result, while weather classification may be useful for indicating the likelihood of extreme conditions, it lacks the precision needed for impact-based surge forecasting.

11.4 Potential interdisciplinary applications

The methodology explored in this study, using weather pattern-based classification to predict storm surge, has broader applications in meteorology and climate impact studies. As summarised by Ireland, Robbins, Neal, Barciela, & Gilbert (2024), this approach is already being used in various forecasting applications. Given its ability to classify large-scale atmospheric configurations, it could be extended to other environmental hazards where atmospheric patterns play a significant role in driving extreme events.

One clear application is in precipitation forecasting. Just as this study attempted to associate weather patterns with storm surge probability, a similar approach could be used to predict heavy rainfall events and associated riverine flooding (Richardson, et al., 2020). Identifying weather patterns linked to extreme precipitation could enhance flood early warning systems by providing probabilistic insights at longer lead times than traditional hydrodynamic models alone.

Another potential application is in forecasting temperature extremes, such as heatwaves and cold spells. Specific weather patterns can be linked to prolonged high or low temperatures. This relationship could be leveraged to improve health risk forecasting, such as anticipating periods of excess mortality linked to extreme temperatures (Huang, et al., 2020).

Beyond direct hazard forecasting, this method could also support renewable energy resource management. Wind and solar energy production are highly sensitive to large-scale atmospheric circulation, and classifying weather patterns could improve energy demand and supply predictions. Wind energy forecasting, for instance, could benefit from identifying circulation patterns that lead to persistent high-wind conditions over offshore wind farms. Similarly, cloud cover patterns associated with large-scale systems could inform solar energy output projections.

12 Conclusion and recommendations

12.1 Conclusions

This study investigated the potential of integrating weather pattern-based storm classification into mid-term storm surge forecasting for the Dutch coast. Two distinct research phases were conducted: Phase I, which evaluated the suitability of Neal et al. (2018) predefined weather patterns, and Phase II, which explored the development of self-clustered weather patterns tailored to Dutch storm surge dynamics.

While Neal's weather patterns offer a structured classification of large-scale atmospheric conditions, the results of this study suggest that their direct application to mid-term storm surge forecasting for the Dutch shore presents several challenges, resulting in limited differentiation between high- and low-surge conditions, especially at the longer lead times. Similarly, the self-clustered weather patterns did not consistently demonstrate significant improvements over the predefined ones, highlighting methodological limitations and potential areas for refinement.

A primary limitation was lead time considerations. The self-clustered patterns tended to describe the weather at the moment of peak surge, rather than the preceding atmospheric conditions that contributed to storm development. As a result, the patterns were not necessarily representative of the conditions that drive surges days in advance, limiting their usefulness for mid-term forecasting.

Additionally, the spatial domain of this study may have constrained predictive performance. Many storm systems affecting the Dutch coast originate over the Atlantic Ocean, beyond the study's selected region. Without accounting for these early-stage developments, the method struggled to capture the evolution of storms at longer lead times.

Despite these limitations, the findings suggest that weather pattern-based classification retains potential as a probabilistic forecasting tool, particularly at shorter lead times. However, the method could also be valuable for longer lead times with further optimisation, calibration, and methodological refinements. Section 12.2 will further explore recommendations for future research, including potential refinements to the methodology, alternative clustering strategies, and expanded datasets that could enhance the predictive skill of weather pattern-based storm surge forecasting.

The remainder of this section addresses the research questions posed in Section 1.6. First, the four sub-questions are answered, followed by the main research question.

What are the key large-scale weather patterns associated with extreme storms impacting the Dutch coast and how do these identified weather patterns correlate with nearshore hydraulic conditions?
Extreme storm surges along the Dutch coast are primarily driven by specific large-scale atmospheric patterns that induce strong onshore winds and low atmospheric pressures over the North Sea. These conditions elevate sea levels and intensify nearshore hydraulic impacts. Key weather patterns include:

- ***Deep low-pressure systems over the North Atlantic and North Sea***: These systems, often associated with extratropical cyclones, generate powerful westerly to north-westerly winds. The prolonged onshore wind stress pushes seawater toward the Dutch coastline, leading to elevated sea levels and increased storm surge potential. For example, Storm Pia (December 2023) developed near Iceland and rapidly deepened as it tracked toward Scandinavia, bringing strong north-westerly winds and significant storm surges along the Dutch coast (Zijderveld, et al., 2024).

- ***Blocking high-pressure systems over Scandinavia or Central Europe***: Such systems can impede the usual west-to-east progression of weather patterns, causing cyclones to stall or redirect over the North Sea. This stagnation can result in sustained wind conditions and prolonged high sea levels along the Dutch coast. For instance, during Storm Xaver (December 2013), a blocking high over southwestern Europe intensified the pressure gradient, leading to prolonged north to north-westerly winds and record-breaking storm surges in the North Sea (Kautz, et al., 2022).

These findings are supported by the exploratory SEAS5 surge event analysis (Section 4.3), which showed that extreme surge events consistently occurred during periods of strong westerly to north-westerly winds, driven by deep low-pressure systems positioned either over the North Sea or just north of Scotland. Furthermore, storm evolution animations (Section 4.4 and Appendix J: Storm evolution in the considered domain) demonstrated that many surge-driving storms originated in the North Atlantic, well outside the study domain, and only entered the model region one or two days before peak surge. This emphasises the importance of considering precursor storm development over a larger spatial domain when attempting to predict storm surges further in advance.

How well do the weather patterns from Neal et al. (2018) **perform in predicting storm surges along the Dutch coast?**

The Neal weather patterns were not able to reliably predict storm surge levels along the Dutch coast. While some patterns showed a higher likelihood of high surge events, the surge distributions were too similar across all patterns, making it difficult to distinguish between those associated with high, moderate, or low surge. This lack of differentiation means that the Neal patterns cannot be used operationally as a stand-alone tool for storm surge forecasting along the Dutch coast.

A key limitation was the broad-scale nature of the Neal patterns, which were originally developed for UK storm risk assessment rather than Dutch coastal flood forecasting. Their spatial coverage and classification criteria did not fully capture the local meteorological and hydrodynamic influences that drive storm surges along the Dutch coast. Since storm surges are highly sensitive to regional wind fields and coastline orientation, the predefined patterns may have been too generalised to accurately resolve the specific conditions that lead to high surges in the Netherlands.

What is the added value of self-clustered weather patterns compared to Neal's predefined weather patterns in mid-term storm surge forecasting?

The self-clustered weather patterns, as defined in this study, did not provide a clear advantage over Neal's predefined patterns in terms of surge predictability in the mid-term lead times. None of the three tested methods produced distinct enough classifications to reliably separate high-surge from low-surge conditions.

Of the three methods:

- Method A (Fixed surge threshold) overrepresented extreme surge events in some locations while underrepresenting them in others, leading to inconsistencies in the clustering results.
- Method B (Maximum Dissimilarity Algorithm, MDA) suffered from severe data limitations due to computational constraints, making its clusters too small to be useful.
- Method C (Stratified sampling) offered the most balanced representation of different surge levels but still struggled to produce clusters that meaningfully distinguished between the severity of storm surge risks.

Method C seems the most reasonable choice, but even this approach did not outperform Neal's patterns in a way that would justify operational implementation in its current form.

However, at shorter lead times, the self-clustered weather patterns demonstrated some improvements over Neal's patterns. The boxplots showing the distribution of surges across the weather (Figures 9-3, 9-12 and 9-18 compared to Figure 6-5) indicate that the range of surge values within each self-clustered pattern was generally less wide, meaning that the classification provided more consistent surge estimates. Additionally, the mean surge values of the self-clustered patterns were more distinct from one another, which aligns with the goal of creating weather patterns that effectively differentiate between high and low surge conditions.

Although the tested methods did not yield clear improvements on the mid-term timescale, self-clustered weather patterns still hold potential advantages over predefined classifications. Custom clustering allows for greater flexibility in adapting to changing storm patterns and could be refined further by incorporating additional variables such as wind stress, storm tracks, surge etc. While this study did not find immediate improvements, future research into optimising self-clustering techniques, potentially with an expanded domain or additional predictors, could provide valuable insights for mid-term storm surge forecasting along the Dutch coast.

What are the potential practical implications and operational utility of integrating weather pattern-based storm classification to mid-term storm forecasting for coastal flood preparedness and response strategies in the Netherlands?

Weather pattern-based storm classification was explored as a complementary tool to existing mid-term storm surge forecasting methods. The results indicate that in its current form, the approach is not yet suitable for operational use, as the generated weather patterns do not consistently provide a clear distinction between high-surge and low-surge events. This limits their immediate applicability for decision-making in storm surge risk management.

However, this does not mean that weather pattern classification has no potential value. The ability to categorise large-scale atmospheric conditions remains a promising avenue for improving situational awareness and probabilistic forecasting. With further refinements, this approach has the potential to enhance mid-term forecast reliability and support decision-making for storm surge preparedness.

At shorter lead times, the weather pattern classification as explored in this study, shows greater potential, particularly when used alongside ensemble forecasting frameworks. Instead of applying the classification blindly at long lead times, it may be more effective when used within an existing SEAS5 forecast. For example, rather than trying to predict a high-surge event 15 days in advance, a more practical approach would be to first generate an SEAS5 ensemble forecast for day 14 and then apply the weather pattern classification to assess surge risk for day 15. The 1-day lead time boxplots indicate that at these shorter timeframes, weather patterns are more distinctly linked to surge probabilities, meaning this approach could help refine probabilistic risk assessments and improve decision-making.

An additional use-case is binary classification, where the goal is to determine whether a high-surge event is likely within a given timeframe. Method A showed promise in this regard, successfully distinguishing between calm and extreme surge days. This suggests that weather pattern classification could serve as a probabilistic screening tool, helping to highlight periods of increased surge risk. Such an approach could be particularly valuable for early warning systems, where forecasters need to assess broad-scale risk levels rather than predict exact surge magnitudes.

At this stage, generating operationally useful weather patterns is not trivial and requires further optimisation before it can be effectively integrated into storm surge forecasting systems. The next section (12.2) provides specific suggestions on how the methodology could be improved to better capture storm surge dynamics and increase its operational value.

To what extent can the integration of weather pattern-based storm classification contribute to improving mid-term forecasting accuracy for extreme storms impacting the Dutch coast?
In its current form, weather pattern-based classification does not significantly improve mid-term storm surge forecasting accuracy. The tested methods did not produce patterns that reliably differentiate between surge levels, limiting their operational usefulness. However, the concept remains promising if further refined.

12.2 Recommendations for future research

The results of this study indicate that while weather pattern-based classification provides a structured approach to analysing storm surge events, its effectiveness in mid-term forecasting remains uncertain. The generated patterns do not consistently align well with high-surge events (as shown in Chapter 9) on the mid-term timescale, highlighting limitations in the current methodology. To improve the applicability of this approach, it is recommended that future research prioritises the following areas as discussed in the following sub-sections.

12.2.1 Priorities for improving weather pattern-based forecasting

To enhance the effectiveness of weather pattern-based classification, several methodological improvements could be explored**:**

I. Introducing lead-time considerations in data selection

One of the most critical limitations of the current approach is that clustering was based on the weather conditions occurring on the high-surge day itself, rather than the atmospheric conditions that led to the surge event. Since storm surge buildup is a multi-day process, future research should shift toward a sequential clustering approach, where weather patterns are classified based on conditions in the days leading up to a surge event rather than the surge day itself.

A lead time of 15 days would likely be a good starting point, as it is long enough to capture the full evolution of a storm system while avoiding arbitrary selection biases in choosing how many preceding days to include. By focusing on the precursor conditions responsible for generating a storm surge rather than the conditions occurring simultaneously with it, this adjustment would make the approach more relevant for mid-term forecasting applications.

This concept is already applied in the boxplot analysis linking maximum daily surge to preceding weather patterns (Figure 5-7). Adopting a similar approach for clustering would help identify causal relationships between weather patterns and surge events.

II. Expanding the clustering domain

The current study only considers a limited domain centred around the North Sea, but many surge-driving storms originate in the North Atlantic before moving into the study region (as shown in Section 4.4). If a storm is already fully developed upon entering the study domain, the clustering process misses the critical precursor conditions that led to its formation.

To better capture the full lifecycle of storms affecting the Dutch coast, the clustering domain should be expanded to include a larger portion of the North Atlantic. This would help resolve the atmospheric conditions that precede extreme storm surges, particularly those linked to deep low-pressure systems tracking from the west.

Additionally, masking land areas within the clustering domain could improve classification quality. Since land-based atmospheric features contribute little to storm surge formation, their inclusion may introduce unnecessary noise into the clustering process. By removing land regions from the classification, the clustering algorithm could focus more directly on oceanic and coastal meteorological patterns, potentially improving pattern distinctiveness.

III. Revisiting data selection criteria: fixed return period vs. fixed surge threshold

Currently, Method A selects high-surge events based on a fixed surge threshold, which does not account for regional differences in typical surge magnitudes. For instance, Delfzijl experiences significantly higher surges than Vlissingen (see Appendix C: Surge return period analysis), meaning that a uniform threshold overrepresents extreme events in some locations and underrepresents them in others.

A more robust alternative is to select events based on a fixed return period rather than a fixed surge threshold. For example, defining storm days based on a 1-year return period at each location would ensure that the clustering includes storms of similar rarity across all sites, improving comparability and better aligning with mid-term forecasting applications for repair and maintenance scheduling.

A related consideration is whether a single set of weather patterns should be derived for the entire Dutch coast or whether separate weather patterns should be defined per location. If weather patterns were generated individually for each site, they would more precisely capture local meteorological influences on storm surge, potentially improving location-specific predictability. However, this would require deriving multiple sets of weather patterns rather than a single set, increasing the workload during the development phase. Future research should explore whether this additional effort leads to meaningful forecast improvements, or if a compromise approach, such as grouping locations with similar storm surge responses into regional subsets, provides a good balance between accuracy and efficiency.

12.2.2 Further methodological improvements

Beyond refining data selection and clustering criteria, alternative techniques and computational enhancements may provide better predictive performance:

IV. Expanding the predictor set

This study relied solely on mean sea level pressure (MSLP) and its gradient as the primary atmospheric predictors for clustering. While MSLP plays a fundamental role in defining large-scale weather patterns, other variables also influence storm surge dynamics. Future research could therefore explore incorporating additional predictors into the clustering process, such as:

- ***Wind stress components*** (zonal and meridional): These determine surface wind forcing, which is a primary driver of storm surge generation.
- ***Sea surface pressure anomalies***: While MSLP was already used, anomalies relative to a climatological mean could highlight pressure deviations that are more relevant for storm activity.
- ***Sea surface temperature***: Warmer waters can contribute to storm development and intensification, affecting the likelihood of strong onshore winds and high surges.
- ***Surge*** itself: While not useful for operational forecasting (as it is the target variable), it could enhance clustering accuracy when generating weather patterns for training purposes.

Additionally, rather than weighting all predictors equally, future research could assign different weights based on their relative importance in driving storm surge events. Given that both pressure and wind effects play an equally significant role in extratropical storm surges in the North Sea (Pugh, 1987), MSLP and wind stress could be given equal weight, while secondary variables such as sea surface temperature might receive lower weighting. These weights could be determined empirically through sensitivity analyses, optimising predictor selection for better classification results.

V. Investigating alternative clustering methods

While k-means clustering is the most commonly used method for weather pattern classification, alternative machine learning approaches may also be tested to assess their ability to capture the complexity of storm surge-related weather patterns. Future research could explore:

- Hierarchical clustering – Allows nested relationships between weather patterns to be analysed, potentially providing a more flexible classification structure (Gueffier, et al., 2024).
- Self-Organising Maps (SOMs) – A neural network-based approach that has shown promise in classifying atmospheric circulation patterns (Doan, Kusaka, Sato, & Chen, 2020).
- Hidden Markov Models (HMMs) – Captures sequential transitions between weather states, making it potentially well-suited for modelling storm evolution (Ailliot, Thompson, & Thomson, 2009).
- Other Machine Learning approaches (Random Forests etc.) – More flexible classification models such as Random Forest, that could provide probabilistic weather pattern assignments for storm surge risk assessment (Bellinghausen, Hünicke, & Zorita, 2024).

Since clustering is ultimately a means of identifying key atmospheric states, testing these alternative methods could provide deeper insights into which weather patterns are most relevant for mid-term surge forecasting.

VI. Expanding data selection for clustering using high-performance computing

A key motivation for using SEAS5 instead of historical datasets (such as ERA5 or EMULATE) is its large dataset size (3.5 million days), which provides a more comprehensive representation of possible weather patterns. However, due to computational limitations in this study, only a small fraction of this dataset was used in clustering; for example, Method A used just 100k days (~275 years). While this is the highest number of days tested in this study, it is only moderately larger than the maximum length of historical datasets available for the Dutch coast (~150 years). Given that SEAS5 has the potential to provide much longer climatological records, it remains unclear whether this sample size is sufficient to fully represent the atmospheric variability present in SEAS5.

To better utilise the full potential of SEAS5, future research should investigate expanding the sample size used in clustering. This could improve the robustness of weather patterns, better capture rare but important surge-driving conditions, and enhance the overall classification of atmospheric states.

One way to achieve this is by leveraging high-performance computing resources, such as DelftBlue. Running the clustering algorithms on an HPC system would allow:

- A significantly larger sample size
- Testing higher values of k (number of clusters)
- Expanding data selection beyond a single location, as in Methods B and C, where storm days were selected based only on surge conditions at IJmuiden.

12.2.3 Operational recommendations for WMCN

Beyond methodological refinements, there are several considerations for how weather-pattern-based classification could be integrated into operational forecasting at WMCN. It is important to emphasise that this approach is not intended to replace numerical storm surge models but rather to serve as a complementary tool. By linking large-scale atmospheric circulation patterns to storm surge probabilities, weather-pattern-based classification could provide additional insights for mid-term risk assessment and decision-making.

A key next step would be to assess whether the self-clustered weather patterns identified in this study could be leveraged for real-time ensemble forecasting applications. The methodology developed in this study was designed with operational use in mind, but due to time constraints, its integration with real-time ECMWF ensemble forecasts was not tested. Future research could explore how well these weather patterns perform when applied to operational ensemble forecasts at different lead times. This would provide insight into whether self-clustered weather patterns can improve probabilistic storm surge predictions and whether adjustments to the classification scheme are needed for practical forecasting applications.

Additionally, since self-clustered weather patterns showed better separation of mean surge levels than Neal's predefined patterns at shorter lead times, their initial implementation could focus on near-term risk assessment rather than extended-range forecasting.

Another operational consideration is whether to derive location-specific weather patterns rather than using a single set of patterns for the entire Dutch coast. While this study classified weather patterns on a national scale, defining patterns per region or station may improve local forecast accuracy. However, this would require testing to determine whether the added effort of deriving multiple location-specific classifications leads to a meaningful improvement in predictive skill.

12.3 Concluding remarks

While the specific self-clustered weather patterns in this study did not yet achieve the desired level of predictive skill, the broader concept of weather pattern-based classification remains promising. Large-scale atmospheric circulation plays a fundamental role in driving storm surges, and with further refinement, this method could still provide valuable insights for mid-term forecasting.

The results of this study should not be seen as a dismissal of the approach but rather as a stepping stone toward improving it. By refining the methodology and integrating it with other forecasting tools, weather pattern-based classification has the potential to become a valuable complement to numerical surge models. Continued research in this direction may unlock new ways to better anticipate storm surge risk at longer lead times, enhancing preparedness and decision-making.

Bibliography

Actueel Hoogtebestand Nederland. (2025). *Dataroom*. Retrieved 02 18, 2025, from AHN: https://www.ahn.nl/dataroom

Aguado, E., & Burt, J. (2001). *Understanding Weather & Climate* (ISBN-10: 0321987306 ed.). Upper Saddle River.

Ailliot, P., Thompson, C., & Thomson, P. (2009). Space–Time Modelling of Precipitation by Using a Hidden Markov Model and Censored Gaussian Distributions. *Journal of the Royal Statistical Society*, 405-426. doi:10.1111/j.1467-9876.2008.00654.x

Alake, R. (2023, 11 24). *Loss Functions in Machine Learning Explained*. Retrieved from datacamp: https://www.datacamp.com/tutorial/loss-function-in-machine-learning?dc_referrer=https%3A%2F%2Fwww.google.com%2F

Allan, R. (2007, 03 01). *EMULATE Mean Sea Level Pressure data set (EMSLP)*. Retrieved 05 13, 2024, from Met Office: https://www.metoffice.gov.uk/hadobs/emslp/

Ansell, T., Jones, P., Allan, R., Lister, D., Parker, D., Brunet, M., . . . Xoplaki, E. (2006). Daily Mean Sea Level Pressure Reconstructions for the European–North Atlantic Region for the Period 1850–2003. *American Meteorological Society*.

Bai, L., & Xu, H. (2022). Accurate storm surge forecasting using the encoder–decoder long short term memory recurrent neural network. *Physics of Fluids*. doi:https://doi.org/10.1063/5.0081858

Bandyopadhyay, S., Maulik, U., & Pakhira, M. (2001). Clustering using simulated annealing with probabilistic redistribution. *International Journal of Pattern Recognition and Artificial Intelligence*, 269-285.

Bell, R., Gray, S., & Jones, O. (2017). North Atlantic storm driving of extreme wave heights in the North Sea. *Journal of Geophysical Research: Oceans*, 122(4). doi:10.1002/2016JC012501

Bellinghausen, K., Hünicke, B., & Zorita, E. (2024). Using Random Forests to Predict Extreme Sea-Levels at the Baltic Coast at Weekly Timescales. *EGU sphere*. doi:https://doi.org/10.5194/egusphere-2024-2222

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* New York: Springer Science + Business Media.

Braverman, M. (2011). *COS597D: Information Theory in Computer Science - Lecture 3.*

Byrne, D., Horsburgh, K., Zachry, B., & Cipollini, P. (2017). Using remotely sensed data to modify wind forcing in operational storm surge forecasting. *Natural Hazards*, Volume 89, pages 275–293.

Camus, P., Mendez, F., Medina, R., & Cofiño, A. (2011). Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, 58(6):453-462. doi:10.1016/j.coastaleng.2011.02.003

Camus, P., Menéndez, M., Méndez, F., Izaguirre, C., Espejo, A., Cánovas, V., . . . Medina, R. (2014). A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research: Oceans*, 7389-7405. doi:10.1002/2014JC010141

Charnock, H. (1955). Wind stress on a water surface. *Royal Meteorological Society, 81*(350), 639-640. doi:https://doi.org/10.1002/qj.49708135027

Chen, S., & Guo, W. (2023). Auto-Encoders in Deep Learning - A Review with New Perspectives. *mathematics*, 1777. doi:https://doi.org/10.3390/math11081777

Chowdhury, R. (2017, 03 18). *Air mass & Frontal depression [Illustration]*. Retrieved from GSP Training Centre (GTC): https://silo.tips/download/air-mass-frontal-depression

Connor, S. (2013, 10 27). *St Jude's Day Storm: The bad news is the long-term forecast's no better*. Retrieved 05 27, 2024, from Independent: https://www.independent.co.uk/news/science/st-jude-s-day-storm-the-bad-news-is-the-longterm-forecast-s-no-better-8907691.html

Costa, W., Idier, D., Rohmer, J., Menendez, M., & Camus , P. (2020). Statistical Prediction of Extreme Storm Surges Based on a Fully Supervised Weather-Type Downscaling Model. *Journal of Marine Science and Engineering*.

de Kraker, A. (2010). Flood events in the southwestern Netherlands and coastal Belgium, 1400-1953. *Hydrological Sciences Journal*, 913-929. doi:https://doi.org/10.1623/hysj.51.5.913

de Vries, H. (2013). *Dealing with extreme storm surges in The Netherlands.* De Bilt: KNMI.

Dekking, F., Kraaikamp, C., Lopuhaä, H., & Meester, L. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How.* London: Springer.

Doan, Q., Kusaka, H., Sato, T., & Chen, F. (2020). A structural self-organizing map (S-SOM) algorithm for weather typing. *Geoscientific Model Development*. doi:https://doi.org/10.5194/gmd-2020-278

DutchNews. (2023, 12 22). *High tides and storm Pia bring flooding to Dutch coastal towns*. Retrieved from DutchNews: https://www.dutchnews.nl/2023/12/high-tides-and-storm-pia-bring-flooding-to-dutch-coastal-towns/

ECMWF. (2018, 01). *ECMWF's new long-range forecasting system SEAS5*. Retrieved 05 13, 2024, from ecmwf: https://www.ecmwf.int/en/newsletter/154/meteorology/ecmwfs-new-long-range-forecasting-system-seas5

ECMWF. (2021). *SEAS5 user guide.*

ECMWF. (2023, 09 27). *What are the changes from ERA-Interim to ERA5 and ERA5-Land?* Retrieved 10 31, 2024, from ECMWF: https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925

ECMWF. (2024, 10 30). *ERA5: data documentation*. Retrieved 10 31, 2024, from ECMWF: https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation

enw | expertisenetwerk waterveiligheid. (2017). *Grondslagen voor hoogwaterbescherming.*

Espejo, A., Camus, P., Losada, I., & Méndez, F. (2014). Spectral Ocean Wave Climate Variability Based on Atmospheric Circulation Patterns. *American Meteorological Society*. doi:DOI: 10.1175/JPO-D-13-0276.1

Etala, P., Saraceno, M., & Echevarría, P. (2015). An investigation of ensemble-based assimilation of satellite altimetry and tide gauge data in storm surge prediction. *Ocean Dynamics*, 435–447.

Feddersen, H., & Andersen, U. (2005). A method for statistical downscaling of seasonal ensemble predictions. *Tellus A: Dynamic Meteorology and Oceanography*, 398-408. doi:10.3402/tellusa.v57i3.14656

Flowerdew, J., Horsburgh, K., & Mylne, K. (2009). Ensemble Forecasting of Storm Surges. *Marine Geodesy*, 91-99. doi:https://doi.org/10.1080/01490410902869151

Fritz. (2023, 09 18). *Understanding the Mathematics behind K-Means Clustering*. Retrieved 01 08, 2025, from fritz: https://fritz.ai/mathematics-behind-k-means-clustering/

Geerse, C. (2020). *Interactie tussen opzet en getij: Tijdsmodellering van de rechte opzet, voortbouwend op ideeën van Horseburgh & Wilson.* Rijkswaterstaat - WVL.

Grönquist, P., Düben, P., Ben-Nun, T., Dryden, N., Lavarini, L., Li, S., & Hoefler, T. (2019). Predicting Weather Uncertainty with Deep Convnets. doi:10.48550/arXiv.1911.00630

Grus, J. (2015). *Data Science from Scratch* (First Edition ed.). United States of America: O'Reilly Media. doi:ISBN: 978-1-491-90142-7

Gueffier, J., Gheusi, F., Lothon, M., Pont, V., Philibert, A., Lohou, F., . . . Vial, A. (2024). Weather regimes and related atmospheric composition at a Pyrenean observatory characterized by hierarchical clustering of a 5-year data set. *European Geosciences Union*. doi:https://doi.org/10.5194/acp-24-287-2024

Gutierrez, D. (2018, 11 23). *Unsupervised Learning: Evaluating Clusters*. Retrieved 01 16, 2025, from ODSC: https://opendatascience.com/unsupervised-learning-evaluating-clusters/

Haenel, V., Gouillart, E., & Varoquaux, G. (2013). *Python Scientific lecture notes.* EuroScipy tutorial team. Retrieved from http://scipy-lectures.github.com

Halava. (2015, 13 5). *Map of the North Sea [Illustration]*. Retrieved from https://en.wikipedia.org/wiki/North_Sea#/media/File:North_Sea_map-en.png

Hayes, A., James, M., & Beer, K. (2024, 08 19). *How Stratified Random Sampling Works, With Examples*. Retrieved from Investopedia: https://www.investopedia.com/terms/stratified_random_sampling.asp

Hegermiller, C., Antolinez, J., Rueda, A., Camus, P., Perez, J., Erikson, L., . . . Mendez, F. (2017). A Multimodal Wave Spectrum–Based Approach for Statistical Downscaling of Local Wave Climate. *American Meteorological Society*. doi:DOI: 10.1175/JPO-D-16-0191.1

HKV. (2024, 04 03). *Over ons*. Retrieved 05 04, 2024, from HKV: https://www.hkv.nl/over-hkv/

Holthuijsen, L. (2007). *Waves in Oceanic and Coastal Waters.* New York: Cambridge University Press.

Holton, J., & Hakim, G. (2013). *An Introduction to Dynamic Meteorology* (5th Edition ed.). Elsevier.

Horsburgh, K., & Wilson, C. (2007). Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research*. doi:10.1029/2006JC004033

Huang, W., Charlton-Perez, A., Lee, R., Neal, R., Sarran, C., & Sun, T. (2020). Weather regimes and patterns associated with temperature-related excess mortality in the UK: a pathway to sub-seasonal risk forecasting. *Environmental Research Letters*. doi:https://doi.org/10.1088/1748-9326/abcbba

Icelandic Met Office. (2015, 11 16). *HARMONIE - numerical weather prediction model*. Retrieved 05 13, 2024, from Icelandic Met Office: https://en.vedur.is/weather/articles/nr/3232

Ireland, L., Robbins, J., Neal, R., Barciela, R., & Gilbert, R. (2024). Generating weather pattern definitions over South Africa suitable for future use in impact-orientated medium-range forecasting. *International Journal of Climatology*. doi:DOI: 10.1002/joc.8396

Jaadi, Z. (2024, 02 23). *Principal Component Analysis (PCA): A Step-by-Step Explanation*. Retrieved 12 02, 2024, from built in: https://builtin.com/data-science/step-step-explanation-principal-component-analysis

Johnson, S., Stockdale, T., Ferranti, L., Balmaseda, M., Molteni, F., Magnusson, L., . . . Monge-Sanz, B. (2019). SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 1087–1117. doi:https://doi.org/10.5194/gmd-12-1087-2019

K. (2023, 03 31). *PCA And It's Underlying Mathematical Principles*. Retrieved 01 08, 2025, from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/09/pca-and-its-underlying-mathematical-principles/

Kautz, L., Martius, O., Pfahl, S., Pinto, J., Ramos, A., Sousa, P., & Woollings, T. (2022). Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review. *Weather and Climate Dynamics*, 305-336. doi:https://doi.org/10.5194/wcd-3-305-2022

Kohno, N., Dube, S., Entel, M., Fakhruddin, S., Greenslade, D., Leroux, M., . . . Thuy, N. (2018). Recent Progress in Storm Surge Forecasting. *Trypical Cyclone Research and Review*, 128-139. doi:https://doi.org/10.6057/2018TCRR02.04

Kraker, A. (2002). *Historic Storms in the North Sea Area, an Assessment of the Storm Data, the Present Position of Research and the Prospects for Future Research* (ISBN : 978-3-642-07744-9 ed.). Climate Development and History of the North Atlantic Realm.

Lee, T. (2006). Neural network prediction of a storm surge. *Ocean Engineering*, 483-494. doi:https://doi.org/10.1016/j.oceaneng.2005.04.012

Li, J., & Nie, B. (2017). Storm surge prediction: present status and future challenges. *IUTAM Symposium on Storm Surge Modelling and Forecasting*, (pp. 3-9). doi:10.1016/j.piutam.2017.09.002

Madsen, K., Høyer, J., Fu, W., & Donlon, C. (2015). Blending of satellite and tide gauge sea level observations and its assimilation in a storm surge model of the North Sea and Baltic Sea. *JGR Oceans*, 6405-6418.

Majewski, P. (2020, 09 04). *Anomaly Detection in Computer Vision with SSIM-AE*. Retrieved from medium: https://medium.com/@majpaw1996/anomaly-detection-in-computer-vision-with-ssim-ae-2d5256ffc06b

Met Office. (2013). *St. Jude's day storm (Cylone Christian).*

Neal, R., Dankers, R., Saulter, A., Lane, A., Millard, J., Robbins, G., & Price, D. (2018). Use of probabilistic medium- to long-range weather-pattern forecasts for identifying periods with an increased likelihood of coastal flooding around the UK. *Meteorological applications*, 534-547. doi:https://doi.org/10.1002/met.1719

Neal, R., Fereday, D., Crocker, R., & Comer, R. (2016). A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteological Applications*, 389–400. doi:10.1002/met.1563

Nieuwhuis, T. (2023). *Modelling the effect of twin storms on dune erosion.*

Northern Vermont University. (n.d.). *Chapter 12 - Mid latitude Cyclones*. Retrieved 05 08, 2024, from Northern Vermont University: https://apollo.nvu.vsc.edu/classes/met130/notes/chapter12/midlat_cyclones_intro.html

Ozer, J., & Legrand, S. (2015). *BEAWARE II : Review of the physical oceanography in the area of the Bonn Agreement.* OD Nature.

Parsons, V. (2017). *Wiley StatsRef: Statistics Reference Online.* John Wiley & Sons, Ltd. doi:10.1002/9781118445112.stat05999.pub2

Port of Rotterdam. (n.d.). *Purpose, missie, visie en strategie*. Retrieved 06 05, 2024, from Port of Rotterdam: https://www.portofrotterdam.com/nl/over-het-havenbedrijf/purpose-missie-visie-en-strategie

Pugh, D. (1987). *Tides, Surges and Mean Sea-Level.* Swindon, UK: John Wiley & Sons.

Ramos-Valle, A., Curchitser, E., Bruyère, C., & McOwen, S. (2021). Implementation of an Artificial Neural Network for Storm Surge Forecasting. *JGR Atmospheres*. doi:10.1029/2020JD033266

Richardson, D., Neal, R., Dankers, R., Mylne, K., Cowling, R., Clements, H., & Millard, J. (2020). Linking weather patterns to regional extreme precipitation for highlighting potential flood events in medium- to long-range forecasts. *Meteorological Applications*. doi:https://doi-org.tudelft.idm.oclc.org/10.1002/met.1931

Rijkoverheid. (2018, 03 27). *Rijkswaterstaat Informatiepunt Water, Verkeer en Leefomgeving (WVL)*. Retrieved 05 04, 2024, from rijksoverheid: https://www.rijksoverheid.nl/contact/contactgids/rijkswaterstaat-water-verkeer-en-leefomgeving-wvl

Rijksoverheid. (2017, 07 28). *Taken van een waterschap*. Retrieved 05 04, 2024, from rijksoverheid: https://www.rijksoverheid.nl/onderwerpen/waterschappen/taken-waterschap

Rijkswaterstaat. (2009). *Water management in the Netherlands [Illustration].*

Rijkswaterstaat. (2019). *Water management in the Netherlands.* Rijkswaterstaat & the Association of Dutch Water Authorities.

Rijkswaterstaat. (2023, 12 21). *Terugblik hoge waterstanden en storm Pia; december 2023*. Retrieved from Rijkswaterstaat: https://www.rijkswaterstaat.nl/nieuws/archief/2023/12/storm-pia

Rijkswaterstaat. (2023, 10 24). *Waterberichtgeving*. Retrieved 05 04, 2024, from rijkswaterstaat: https://www.rijkswaterstaat.nl/water/waterbeheer/watermanagementcentrum-nederland/waterberichtgeving

Rijkswaterstaat. (2024, 04 29). *Crisisadvisering*. Retrieved 05 04, 2024, from rijkswaterstaat: https://www.rijkswaterstaat.nl/water/waterbeheer/watermanagementcentrum-nederland/crisisadvisering

Rozas Rojas, C. (2017). *Assessing the impact of climate change on longshore sediment transport along the central Dutch coast using statistical downscaling.* Delft: Delft University of Technology.

Sampaio, C. (2023, 11 17). *Definitive Guide to K-Means Clustering with Scikit-Learn*. Retrieved 01 16, 2025, from StackAbuse: https://stackabuse.com/k-means-clustering-with-scikit-learn/

Schoof, J. (2013). Statistical Downscaling in Climatology. *Geography Compass*, 249-265. doi:10.1111/gec3.12036

Schwierz, C., Köllner-Heck, P., Mutter, E., Bresch, D., Vidale, P., Wild, M., & Schär, C. (2010). Modelling European winter wind storm losses in current and future climate. *Climatic Change*, 101:485–514. doi:10.1007/s10584-009-9712-1

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 379-423. doi:10.1002/j.1538-7305.1948.tb01338.x

Sharma, P. (2024, 05 21). *The Ultimate Guide to K-Means Clustering: Definition, Methods and Applications*. Retrieved 05 27, 2024, from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 1929-1958.

Steenhuis, M. (2015). *water & heritage: material, conceptual and spiritual connections.* Leiden: Sidestone Press.

Sterl, A., van den Brink, H., de Vries, H., Haarsma, R., & van Meijgaard, E. (2009). An ensemble study of extreme storm surge related water levels in the North Sea in a changing climate. *Ocean Science*, 5, 369–378.

Stolk, A., & Dillingh, D. (1989). *A short review of the Dutch coast.* Commission of the European Communities.

Stolte, W., Baart, F., Muis, S., Hijma, M., Taal, M., Le Bars, D., & Drijfhout, S. (2023). *Zeespiegelmonitor.* Deltares.

Sumihar, J., & Muis, S. (2024). *Exploration long term predictions of wind and air pressure in the North Sea.* Delft: Deltares.

Sünderman, J., & Pohlmann, T. (2011). A brief analysis of North Sea physics. *Oceanologia*, 663-689. doi:doi:10.5697/oc.53-3.663

Toyoda, M., Fukui, N., Miyashita, T., Shimura, T., & Mori, N. (2021). Uncertainty of storm surge forecast using integrated atmospheric and storm surge model: a case study on Typhoon Haishen 2020. *Coastal Engineering Journal*, 135-150.

TU Delft, SLIMM Lab & Aidrolab. (2024). *CEGM2003 - Data Science and Artificial Intelligence for Engineers*. Retrieved from https://interactivetextbooks.citg.tudelft.nl/dsaie/intro.html

van Aalst, J. W. (2019, 10 22). *De 21 waterschappen [illustration].* Retrieved from wikipedia: https://nl.wikipedia.org/wiki/Lijst_van_Nederlandse_waterschappen#/media/Bestand:2019-Waterschap-prov-1200.png van den Brink, H. (2020). *Het gebruik van de ECMWF seizoensverwachtingen voor het berekenen van de klimatologie van extreme waterstanden langs de Nederlandse kust.* De Bilt: Koninklijk Nederlands Meteorologisch Instituut.

van den Brink, M. (2020). *Guardians of Public Value: How Public Organisations Become and Remain Institutions* (ISBN 978-3-030-51700-7 ed.). palgrave macmillan.

van der Spek, A., van der Werf, J., Oost, A., Vermaas, T., Grasmeijer, B., & Schrijvershof, R. (2022). The lower shoreface of the Dutch coast – An overview. *Ocean and Coastal Mangament*. doi:https://doi.org/10.1016/j.ocecoaman.2022.106367

Verlaan, M., Zijderveld, A., de Vries, H., & Kroos, J. (2005). Operational storm surge forecasting in the Netherlands: developments in the last decade. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 1441-1453. doi:10.1098/rsta.2005.1578

von Storch, H. (2014). Storm surges: Phenomena, forecasting and scenarios of change. *23rd International Congress of Theoretical and Applied Mechanics*, (pp. 356 – 362). doi:doi: 10.1016/j.piutam.2014.01.030

Wadey, M., Haigh, I., Nicholls, R., Brown, J., Horsburgh, K., Carroll, B., . . . Bradshaw, E. (2015). A comparison of the 31 January - 1 Februari 1953 and 5-6 December 2013 coastal flood events around the UK. *Frontiers in Marine Science*.

Wai-hung, L. (2010, 12). *Geostrophic Wind*. Retrieved from Hong Kong Obervatory: https://www.hko.gov.hk/en/education/weather/meteorology-basics/00010-geostrophic-wind.html

Willett, P. (1999). Dissimilarity-Based Algorithms For Selecting Structurally Diverse Sets Of Compounds. *Journal of Computational Biology*, 447-457.

WMCN. (2023). *Landelijk Draaiboek Hoogwater en Overstromingsdreiging.* Lelystad: Watermanagementcentrum Nederland (WMCN).

Woth, K. (2005). *Regionalization of global climate change scenarios: An ensemble study of possible changes in the North Sea storm surge statistics.* Hamburg: Universität Hamburg.

Xie, W., Xu, G., Zhang, H., & Dong, C. (2023). Developing a deep learning-based storm surge forecasting model. *Ocean Modelling*. doi:https://doi.org/10.1016/j.ocemod.2023.102179

Zijderveld, A., Netel, S., Verboeket, R., Driebergen, J., van Schaik, M., & van Galen, L. (2024). *Stormvloedrapport SR101.* Lelystad: Watermanagementcentrum Nederland (WMCN).

Zijderveld, A., Verboeket, R., Bosma, B., & IJpelaar, R. (2022). *Stormvloedrapport SR100.* Watermanagementcentrum Nederland.

Appendix A: Autoencoder

A.1 Introduction to autoencoders

An autoencoder is a type of artificial neural network designed to learn a compressed representation of input data. It has two main parts:

- ***Encoder*** - Compresses the input data into a smaller, "latent" space representation. This compression process captures essential features while discarding unnecessary details.
- ***Decoder*** - Reconstructs the original data from this compressed form, aiming to make the output as similar to the input as possible.

For this study, autoencoders are used to learn the fundamental patterns in SEAS5 mean sea level pressure (MSLP) data and compare these patterns to the predefined Neal weather patterns.

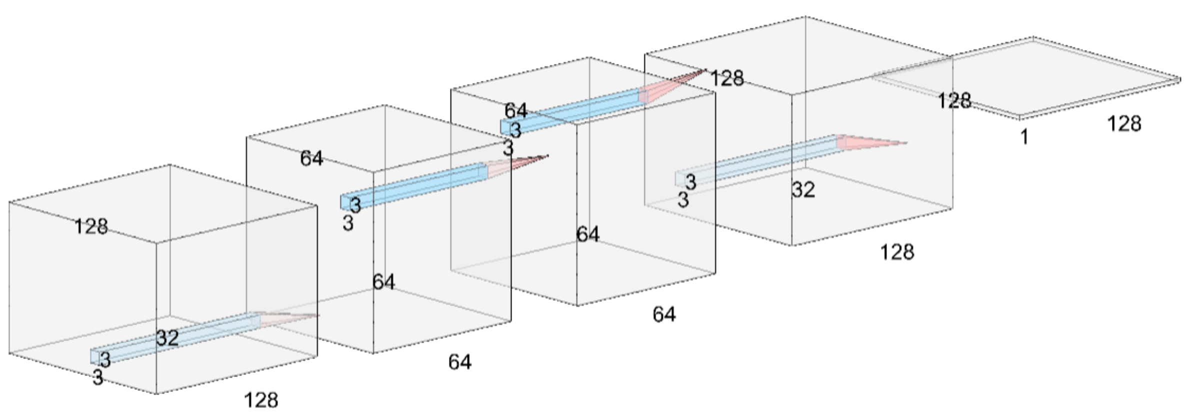
A.2 Why use an autoencoder for pattern matching?

The main advantage of using an autoencoder is its ability to capture complex spatial patterns. By training the autoencoder on SEAS5 data, a so-called "latent space" is created that represents key atmospheric features. The trained encoder can then transform both SEAS5 and Neal patterns into this latent space, allowing for a meaningful comparison of the patterns.

A.3 Autoencoder architectures used

Three different autoencoder architectures were tested, each with increasing complexity. The differences in complexity allowed exploration of how varying network depths and features affected the autoencoder's ability to capture essential atmospheric patterns. The architectures are illustrated in Figure A-1, with the following specifications:

- ***Simple architecture*** - Compresses the data to a low-dimensional representation and reconstructs it with minimal complexity.
- ***Intermediate architecture*** - Increases the depth of the encoder and decoder, with additional filters to capture more details.
- ***Complex architecture*** - Uses a much deeper encoder and decoder, compressing data further and capturing finer spatial details through additional layers.



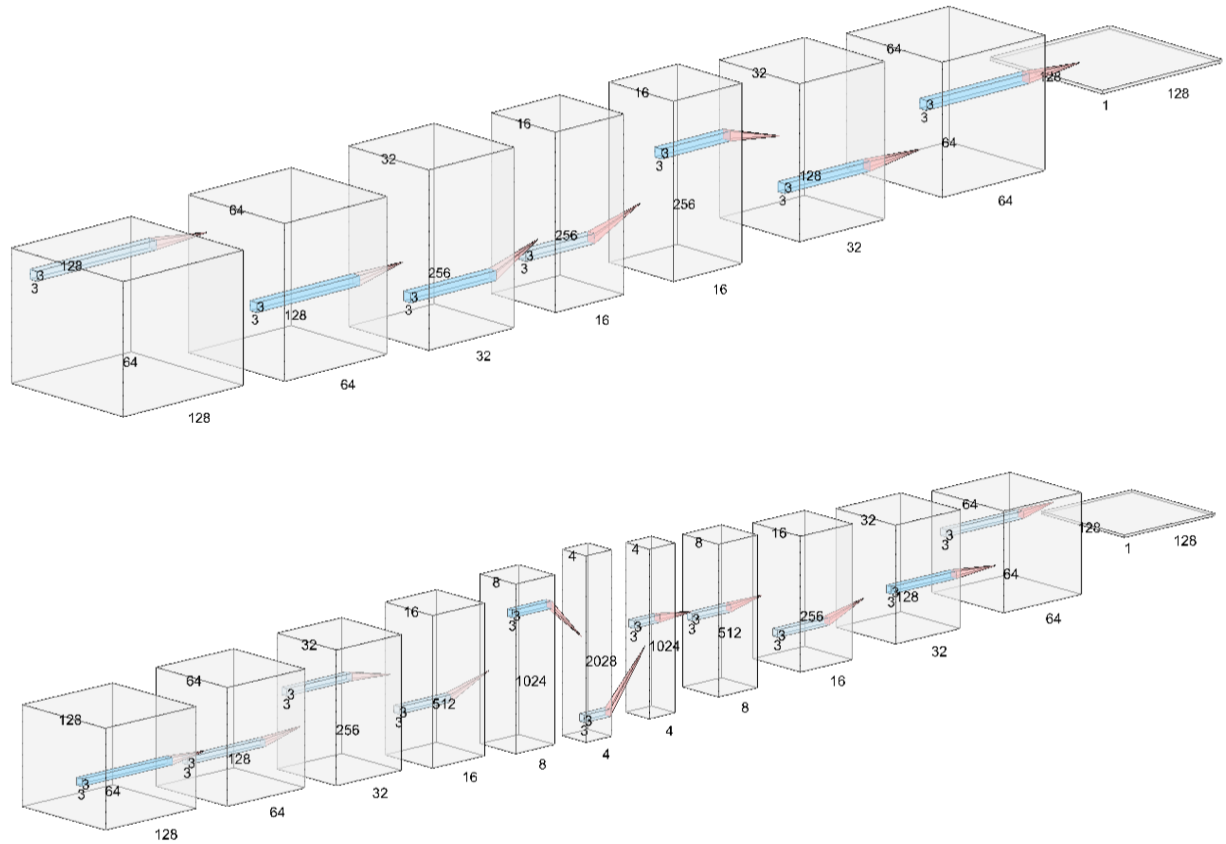


Figure A-1: Three autoencoder architectures used for matching SEAS5 MSLP fields to Neal patterns, displayed in increasing complexity from top to bottom. Top: simple architecture, middle: intermediate architecture, bottom: complex architecture.

A.4 Key hyperparameters in the autoencoder model

Hyperparameters are settings that define the structure and operation of the model. Adjusting these can significantly impact how well the autoencoder learns to represent the data. Here are some key hyperparameters used in this study, explained for those unfamiliar with deep learning concepts (Chen & Guo, 2023):

- ***Hidden layers -*** Hidden layers add depth to the network, allowing it to learn more complex patterns. Adding more layers can increase the model's ability to capture detailed structures but also raises the risk of "overfitting" (fitting too closely to the training data, reducing generalisation).
- ***Neurons per layer -*** Each layer consists of neurons, which process pieces of information. More neurons allow the model to capture finer details but also increase the risk of overfitting.
- ***Latent space size -*** The latent space is the compressed representation of the data. A smaller latent space forces stronger compression, discarding more details, while a larger latent space retains more information but may include unnecessary details.
- ***Activation function -*** Activation functions introduce non-linear patterns, helping the model capture complex relationships. Common choices include ReLU (Rectified Linear Unit), sigmoid, and tanh. Each function has unique properties that affect how the model processes information (Chen & Guo, 2023).

- ***Loss function -*** This function measures the difference between the input and the reconstructed output, guiding the model to improve its accuracy. The loss function tells the autoencoder how close its reconstruction is to the original data. Two loss functions were tested:
    - ***Mean Squared Error (MSE) -*** Focuses on reducing pixel-level differences between the input and output images (Alake, 2023).
    - ***Custom hybrid loss -*** Combines MSE with Structural Similarity Index Measure (SSIM), which evaluates structural similarity. SSIM is beneficial for retaining visual structure, as it assesses contrast, luminance, and structural information, reflecting how humans perceive visual details (Majewski, 2020).

A.5 Model regularisation techniques: dropout and batch normalisation

To prevent overfitting and help the model generalise better, two regularisation techniques were used:

- ***Dropout -*** During training, dropout randomly deactivates certain neurons in each layer, preventing the model from becoming too reliant on specific neurons and thus helping it generalise better (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).
- ***Batch normalisation -*** This technique normalises the output of each convolutional layer, stabilising the learning process and accelerating convergence. By maintaining consistent input distributions across layers, batch normalisation improves model efficiency and robustness (Chen & Guo, 2023).

These techniques contribute to faster convergence and better generalisation, helping the model learn more stable representations of SEAS5 data.

A.6 Training configuration and hyperparameter comparison

Each autoencoder architecture was trained with different configurations to optimise learning. Below is a comparison of the three main versions:

Table A-0-1: Comparison autoencoder architecture

Aspect	Simple	Intermediate	Complex
Architecture complexity	Simple encoder-decoder	Deeper with regularisation	Increased depth and more filters
Encoder layers	2 layers; 32, 64 filters	3 layers: 64, 128, 256 filters	4 layers: 64, 128, 256, 512 filters
Decoder layers	2 layers: 64, 32 filters	3 layers: 256, 128, 64 filters	4 layers: 512, 256, 128, 64 filters
Dropout layers	None	3 dropout layers; rate of 0.3	3 dropout layers; rate of 0.25
Batch normalisation	None	Applied after each layer	Applied after each layer
Learning rate	1e-4	1e-4 with decay	1e-5 with decay
Loss function	MSE	Custom SSIM with MSE	Custom SSIM with MSE
Data augmentation	None	Yes	Yes, with more variations
Epochs	10	100	100
Early stopping	No	Yes, patience 10 epochs	Yes, patience 5 epochs

These settings control how the model learns and help determine the final quality of pattern recognition.

Appendix B: Theory explained

This appendix provides a detailed explanation of key theoretical concepts relevant to this study. These explanations are intended to assist readers who may require additional context or a refresher on the foundational principles. The topics covered include: Matching metrics (B.1), K-means clustering (B.2) and Principal Component Analysis (B.3).

B.1 Matching algorithms: Euclidean distance, RMSE and Pearson correlation

In Section 5.2 of the main report, various matching algorithms were employed to compare SEAS5 mean sea level pressure (MSLP) fields with weather patterns from the Neal dataset. The methods used included Euclidean distance, Root Mean Square Error (RMSE), and Pearson correlation. These metrics provide distinct ways to quantify the similarity or difference between the SEAS5 fields and the Neal dataset weather patterns. Additionally, an autoencoder model was utilised for matching purposes; further details regarding the autoencoder can be found in Appendix A: Autoencoder

Euclidean distance

The first approach used to match SEAS5 mean sea level pressure (MSLP) fields to Neal's weather patterns is based on Euclidean distance. Neal's patterns were created using k-means clustering, which itself minimises Euclidean distance within clusters. Therefore, using Euclidean distance for the matching process is consistent with the methodology Neal used to define these patterns, making it a logical choice for comparison.

Each SEAS5 MSLP field is matched to the closest Neal pattern by calculating the Euclidean distance between the MSLP field and each of Neal's 30 predefined patterns. For each time step in the SEAS5 dataset, the Euclidean distance is computed as follows:

$$d = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

where d represent the Euclidean distance, p_i are the SEAS5 MSLP values, q_i are the corresponding values in a Neal pattern, and n is the total number of grid points. The pattern with the smallest Euclidean distance is identified as the best match, representing the closest resemblance to SEAS5's atmospheric state at that time.

Root Mean Squared Error (RMSE)

The second used approach to match SEAS5 MSLP fields with Neal's weather patterns is based on Root Mean Squared Error (RMSE). RMSE is a commonly used measure of the average magnitude of error between two sets of values, providing an indication of similarity between datasets by quantifying the average deviation. Using RMSE allows for a comparison between SEAS5 and Neal patterns in a way that highlights the overall fit between their respective pressure values across the grid.

For each time step in the SEAS5 dataset, the RMSE is calculated between the SEAS5 MSLP field and each of Neal's 30 predefined weather patterns. RMSE is computed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - q_i)^2}$$

where $RMSE$ represents the Root Mean Squared Error p_i are the SEAS5 MSLP values, q_i are the corresponding values in a Neal pattern, and n is the total number of grid points. The Neal pattern with the lowest RMSE value is considered the best match, representing the closest fit to SEAS5's atmospheric state for that time step.

Pearson correlation

The third approach used to match SEAS5 MSLP fields to Neal's weather patterns is based on Pearson correlation. Pearson correlation measures the linear relationship between two datasets, providing a value between -1 and 1. A correlation close to 1 indicates a strong positive relationship, where the pressure values in SEAS5 align closely with those in Neal's patterns. Conversely, a correlation near -1 suggests an inverse relationship, while a value around 0 indicates no linear relationship.

For each time step in the SEAS5 dataset, Pearson correlation is calculated between the SEAS5 MSLP field and each of Neal's 30 predefined weather patterns. The Pearson correlation coefficient is computed as follows:

$$r = \frac{\sum_{i=1}^{n}(p_i - \bar{p})\,(q_i - \bar{q})}{\sqrt{\sum_{i=1}^{n}(p_i - \bar{p})^2}\sqrt{\left(\sum_{i=1}^{n}(q_i - \bar{q})^2\right)}}$$

where r represents the Pearson correlation coefficient, p_i are the SEAS5 MSLP values, $q_i$are the corresponding values in a Neal pattern, $\bar{p}$ and $\bar{q}$ are the mean values of the SEAS5 and Neal data, respectively, and n is the total number of grid points. The Neal pattern with the highest correlation (closest to 1) is identified as the best match, representing the closest linear relationship to SEAS5's atmospheric state for that time step.

B.2 K-means clustering

K-means clustering is one of the most widely used unsupervised machine learning algorithms for partitioning a dataset into distinct groups or clusters. It operates iteratively to assign each data point to one of k predefined clusters based on its features (Grus, 2015). This section explains the four fundamental steps of the k-means algorithm: initialisation, cluster assignment, centroid update, and repeat-until-convergence (Sharma, 2024). The accompanying figures illustrate these steps.

Initialisation

The first step in K-means clustering is the initialisation of cluster centroids. k centroids are chosen randomly within the feature space of the dataset. These centroids serve as the starting points for the clustering process. The initial placement of centroids can have a significant impact on the algorithm's outcome and can influence the final clusters' quality. Figure B-1 illustrates the random initialisation of centroids for a sample dataset.

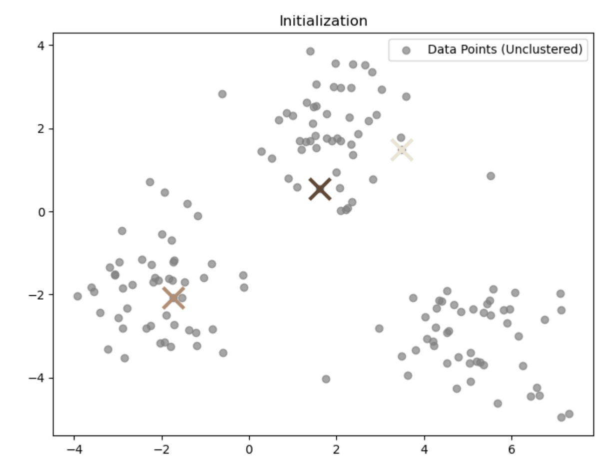


Figure B-1: Random placement of initial centroids among unclustered data points to start the k-means process.

Cluster assignment

In the cluster assignment step, each data point x_i in the dataset is assigned to the cluster whose centroid is closest to it. The closest cluster is determined using a predefined distance metric; often the Euclidean distance is chosen. Mathematically, the cluster j assigned to a data point x_i is determined by (Fritz, 2023):

$$j = arg\,\min_k \|x_i - \mu_k\|^2$$

Here, μ_k represents the centroid of the k-th cluster and $\|x_i - \mu_k\|^2$ is the Euclidean distance between x_i and μ_k. This step effectively partitions the dataset into k clusters based on proximity to the centroids. Figure B-2 demonstrates the result of the cluster assignment process for the initial iteration.

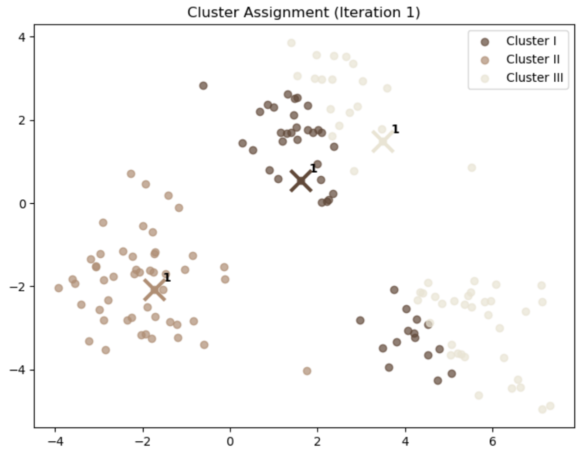


Figure B-2: Data points are assigned to their closest cluster centroids based on the Euclidean distance, forming the initial clusters.

Update of centroids

After assigning data points to clusters, the centroids of each cluster are updated. The new centroid of a cluster is calculated as the mean of all data points assigned to that cluster. The mathematical expression for updating the centroid μ_k of cluster k is:

$$\mu_k = \frac{1}{N_k} \sum_{x_i \in C_k} x_i$$

where:

- C_k – is the set of points assigned to cluster k
- N_k – is the number of points in C_k
- μ_k – is the new centroid of cluster k

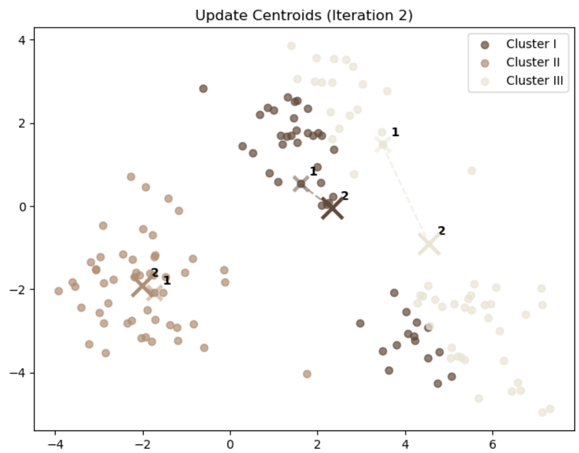


Figure B-3: Cluster centroids are recalculated based on the mean position of the data points within each cluster.

This computation ensures that the centroid reflects the central location of the points in its cluster. Figure B-3 illustrates the updated centroids after the first iteration.

Iterate until convergence

The cluster assignment and centroid update steps are repeated iteratively until the algorithm converges. Convergence is typically defined as one of the following definitions (Sharma, 2024) :

i. when the centroids no longer change significantly
ii. when the cluster assignments stabilise
iii. when a pre-defined maximum number of iterations is reached

This iterative process ensures that the clusters are refined progressively, leading to a locally optimal partition of the dataset. The plots in Figure B-4 illustrate the iterative updates of centroids and cluster assignments until convergence is achieved. In this case, convergence is defined as the state where the cluster assignments stabilise.

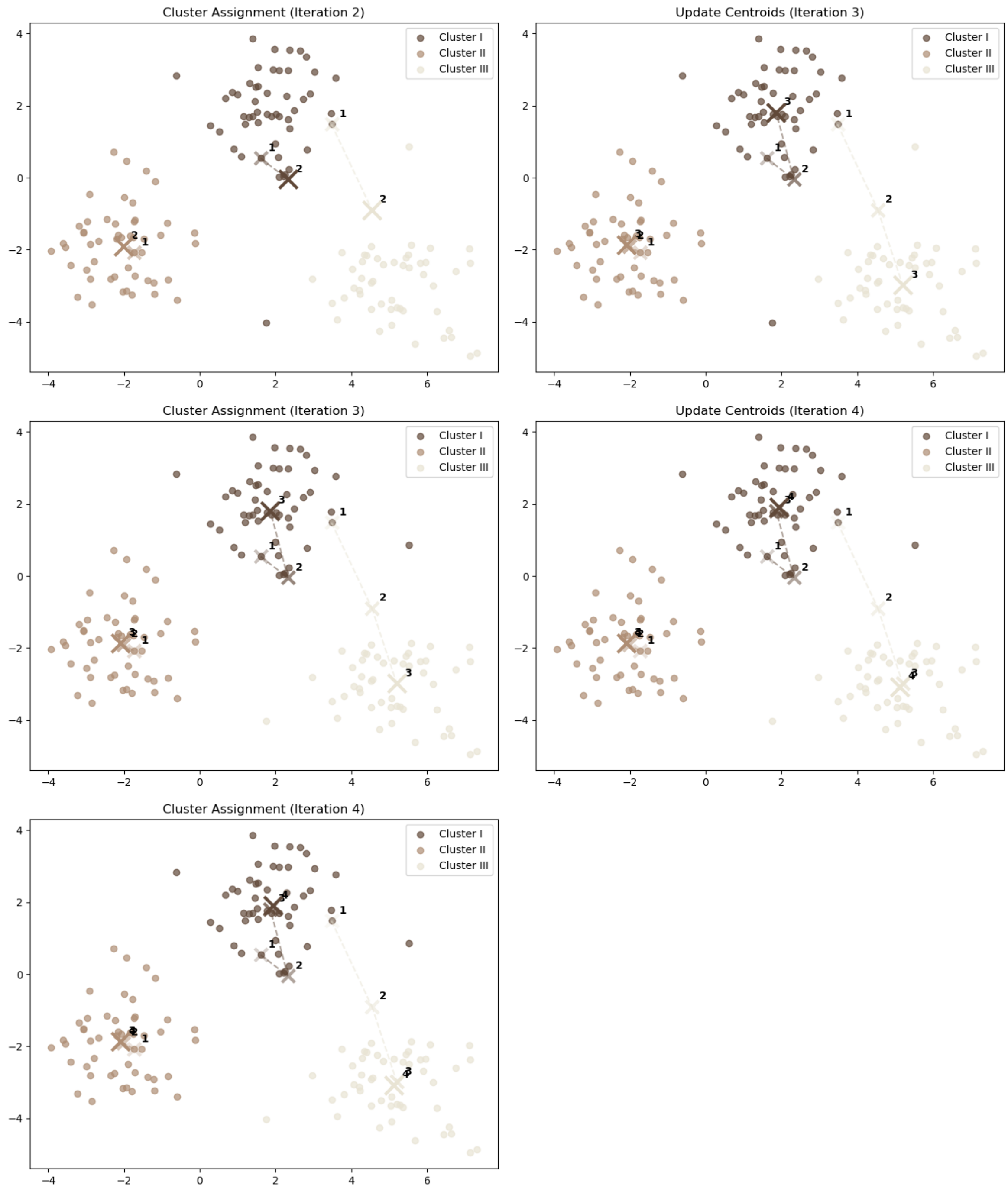


Figure B-4: Subsequent iterations demonstrate the refinement of clusters through iterative re-assignment of data points and updates to centroids until convergence is achieved. Convergence in this case is marked by stabilised cluster assignment.

Determining the optimal number of clusters k

Selecting an appropriate number of clusters, k, is a critical step in applying the k-means clustering algorithm. An inadequate choice of k can lead to underfitting or overfitting, potentially reducing the interpretability or effectiveness of the clustering. Various techniques are available to guide the selection of k, including the elbow method, silhouette score, and statistical metrics like the within-cluster sum of squares (WCSS) and between-cluster sum of squares (BCSS), which were utilised in this project. These measures are conceptually explained below.

1. **Within-Cluster Sum of Squares (WCSS)**

WCSS is a metric that quantifies the compactness of clusters by measuring how tightly data points are grouped around their respective centroids. Specifically, WCSS is the sum of squared distances between each data point and the centroid of its cluster. It is mathematically expressed as (Sampaio, 2023):

$$WCSS = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

where:

- C_k – is the set of data points in cluster k
- x_i – is an individual data point
- μ_k – is the centroid of cluster k

As k increases, WCSS decreases because more clusters result in tighter groupings of points. However, diminishing returns are observed beyond a certain value of k, where adding more clusters provides little improvement in compactness. This forms the basis of the elbow method, where the "elbow" in the WCSS curve indicates the optimal k.

2. **Between-Cluster Sum of Squares (BCSS)**

BCSS measures the separation between clusters by quantifying the dispersion of cluster centroids relative to one another. It is computed as the average squared distance between all pairs of centroids and is given by (Sampaio, 2023):

$$BCSS = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=i+1}^{K} \|\mu_i - \mu_j\|^2$$

where:

- N – is the total number of unique centroid pairs
- μ_i and μ_j – are the centroids of clusters i and j, respectively

Higher BCSS values indicate better-separated clusters, which is desirable in clustering tasks. As k increases, BCSS initially grows because the clusters become more distinct, but it stabilises when adding more clusters no longer increases the separation significantly (Gutierrez, 2018).

The selection of k is guided by examining the behavior of WCSS and BCSS together. The ideal k is chosen to balance these two objectives:

- Minimise WCSS to ensure compact and cohesive clusters
- Maximise BCSS to ensure well-separated and distinct clusters

Figure B-5 illustrates the concepts of Within-Cluster Sum of Squares (WCSS) and Between-Cluster Sum of Squares (BCSS) using a two-cluster example. The left plot demonstrates WCSS by showing the intra-cluster variance, where the lines represent the distances between data points and their respective cluster centroids. This variance reflects the compactness of each cluster. The right plot visualises BCSS by showing the distance between the centroids of the two clusters, represented by the dashed line. This distance indicates the separation between clusters.

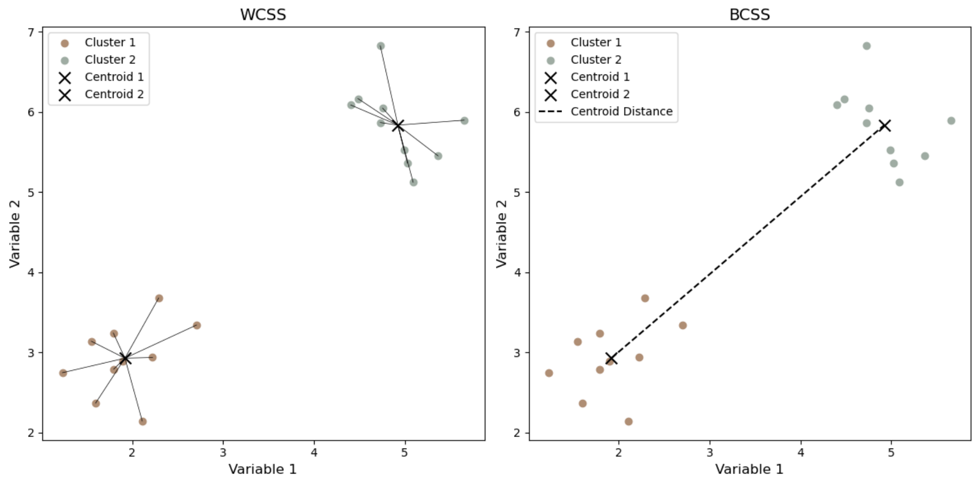


Figure B-5: Visualisation of intra-cluster variance (WCSS) and inter-cluster distance (BCSS) in a two-cluster example.

By plotting WCSS and BCSS against k, the elbow method is applied. The optimal k corresponds to the point where WCSS begins to flatten (indicating diminishing returns), and BCSS stabilises, showing that further increases in k do not improve separation significantly.

B.3 Principle Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique widely used to simplify high-dimensional datasets while retaining as much of the original information as possible. PCA achieves this by identifying new orthogonal directions, called principal components, along which the data varies the most. These principal components allow for a reduced representation of the data that is computationally efficient and easier to interpret (Grus, 2015; Haenel, Gouillart, & Varoquaux, 2013).

Reasons for applying PCA in this study

When clustering data using k-means, the input data must first be represented in a standardised feature matrix $\boldsymbol{X_{std}}$. In this study, the predictors used are the mean sea level pressure (MSLP) and its gradient. The data has a fine spatial resolution, with 73x107 grid points, amounting to 7,811 datapoints per predictor. For a single day, this results in a total of 15,622 datapoints (features). With many days included in the dataset, the original feature matrix becomes extremely large and computationally expensive to process:

$$\boldsymbol{X_{std}} = \begin{bmatrix} Day\ 1 & \longrightarrow & [mslp\_1, \dots, mslp\_n, gradient\_1, \dots, gradient\_n] \\ Day\ 2 & \longrightarrow & [mslp\_1, \dots, mslp\_n, gradient\_1, \dots, gradient\_n] \\ \vdots & \vdots & \vdots \\ Day\ T & \longrightarrow & [mslp\_1, \dots, mslp\_n, gradient\_1, \dots, gradient\_n] \end{bmatrix}$$

This original high-dimensional feature matrix poses two main challenges:

1. ***Dimensionality reduction*** - The high number of features makes clustering computationally intensive. PCA reduces the dimensionality of $\boldsymbol{X_{std}}$ by transforming it into a smaller set of principal components (PCs) that still capture the majority of the dataset's variance. The transformation is represented as:

$$\boldsymbol{X_{std}} \xrightarrow{PCA} \boldsymbol{X_{std,reduced}}$$

where the reduced matrix $\boldsymbol{X_{std,reduced}}$ has the same rows (samples) but a much smaller number of columns (principal components):

$$\boldsymbol{X_{std,reduced}} = \begin{bmatrix} Day\ 1 & \longrightarrow & [PC\_1, \dots PC\_N] \\ Day\ 2 & \longrightarrow & [PC\_1, \dots PC\_N] \\ \vdots & \vdots & \vdots \\ Day\ T & \longrightarrow & [PC\_1, \dots PC\_N] \end{bmatrix}$$

2. ***Noise reduction*** - With high-dimensional data, the Euclidean distances used in k-means clustering can become less meaningful. This phenomenon, known as the "curse of dimensionality, makes it difficult to distinguish between clusters, as the distances are overwhelmed by noise from irrelevant or redundant features " (Fritz, 2023; Grus, 2015). By retaining only the most significant principal components, PCA effectively filters out noise, ensuring that the clustering results are more meaningful.

Thus, PCA is applied before clustering to prepare a reduced feature matrix where the dimensionality is smaller, and the data is denoised. This reduced matrix is subsequently used as input for the k-means algorithm, ensuring more reliable and efficient clustering.

More details on k-means clustering may be found in section B.1 of this appendix. The rest of this section explains the mathematical steps involved in applying PCA and gives a simplified visual example of PCA.

PCA - mathematically explained
The PCA process involves several steps (K, 2023):

I. Standardisation of the dataset

PCA begins by standardising the dataset so that features with larger scales do not dominate the results. Each feature is centred by subtracting its mean and scaled to have unit variance. For a dataset $\boldsymbol{X}$ with n samples and p features, the standardised dataset is:

$$\boldsymbol{X_{std}} = \frac{\boldsymbol{X} - \mu}{\sigma}$$

where μ is the mean vector, and σ is the standard deviation vector of the features.

II. Compute the covariance matrix

The covariance matrix of the standardised data is computed to capture the relationships between different features. It is defined as:

$$\boldsymbol{C} = \frac{1}{n-1} \boldsymbol{X_{std}^{\top}} \boldsymbol{X_{std}}$$

The covariance matrix is a square symmetric matrix where each element represents the covariance between two features.

III. Calculate eigenvectors and eigenvalues

Next, eigenvectors and eigenvalues of the covariance matrix are computed. The eigenvectors represent the principal components, which define the new axes of the transformed dataset, and the eigenvalues indicate the amount of variance captured by each principal component.

$$\boldsymbol{Cv} = \lambda \boldsymbol{v}$$

Here, $\boldsymbol{v}$ is an eigenvector (principal component), and λ is its corresponding eigenvalue, representing the variance explained by that component.

IV. Sort and select principal components

The eigenvalues are sorted in descending order, and the top m eigenvectors corresponding to the largest eigenvalues are selected. This selection determines the dimensionality of the reduced dataset.

The explained variance indicates the proportion of the dataset's total variance captured by each principal component and is calculated as:

$$Explained\ Variance\ Ratio = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

To decide how many principal components to retain (e.g. to determine m) , a scree plot is often used. A scree plot displays the eigenvalues (or explained variance) in descending order against the principal component index.

In this study, a cumulative scree plot was used to determine the number of principal components to retain. Principal components were selected such that their cumulative explained variance reached 99%. This approach ensures that nearly all the variability in the dataset is preserved while reducing the dimensionality.

V. Transform the data

The data is projected onto the selected principal components to obtain the reduced representation:

$$\boldsymbol{Z} = \boldsymbol{X_{std}W}$$

Here, $\boldsymbol{W}$ is the matrix of selected eigenvectors, and $\boldsymbol{Z}$ is reduced dataset.

PCA - visually explained

This sub-section explains the principles of PCA through a visual demonstration using a 3D example. PCA is commonly applied to datasets in high-dimensional spaces with many more features (axes). However, as such spaces cannot be visualised directly, a simplified 3D example is used to provide an intuitive understanding of how PCA works.

Figure B-6 shows a dummy dataset in its original form, represented in a 3D feature space. Each axis corresponds to one of the original features (e.g., X, Y, and Z). The data points are scattered across this space, and their distribution reflects the variance captured by each feature.

Overlaying this dataset are the principal component axes ($PC1$, $PC2$, and $PC3$), which are computed as the directions that capture the greatest variance in the data. These principal components are orthogonal to each other and are ranked in order of the variance they explain. Thus, $PC1$ is the direction of maximum variance; $PC2$ is the direction orthogonal to $PC1$ that explains the second most variance and $PC3$ is the direction orthogonal to both $PC1$ and $PC2$, explaining the remaining variance. This visualisation demonstrates how PCA identifies new axes that better represent the structure of the data.

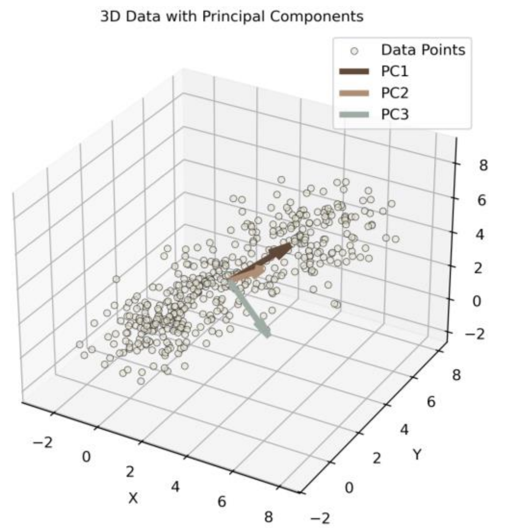


Figure B-6: Visualisation of a dummy dataset in 3D, with principal components overlaid to show the directions of maximum variance.

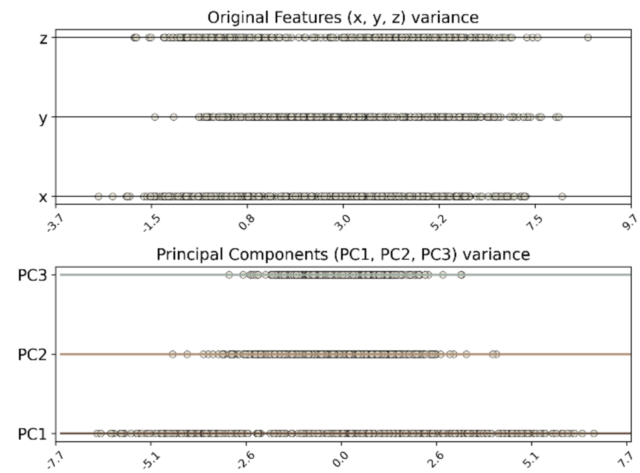


Figure B-7: Comparison of variance in the original features (X, Y, Z) and the PCA latent space (PC1, PC2, PC3).

The top plot of Figure B-7 projects the data onto the original feature axes (X, Y, and Z). Each data point's position along these axes reflects its values in the original feature space. This plot reveals the relative contributions of the features to the dataset's overall variance, as well as potential redundancy, where multiple features may be capturing similar information. PCA seeks to reduce such redundancy by re-expressing the dataset in terms of principal components, which better summarise the variance in the data.

The bottom plot of Figure B-7 plot shows the same data projected onto the principal component axes ($PC1$, $PC2$, and $PC3$). In this new space $PC1$ captures the majority of the variance, as evidenced by the wider spread of the data along this axis. The $PC2$, and $PC3$ components capture progressively less variance. This transformation reveals the underlying structure of the data, simplifying the dataset while retaining most of its variability. By reducing the dataset to just the first few principal components (e.g., $PC1$ and $PC2$), PCA achieves dimensionality reduction without significant loss of information.

B.4 Computing wind speed from wind stress using the Charnock method

The wind speed was derived from the wind stress fields provided in SEAS5, using the Charnock method, with an additional correction factor applied for bias correction in the WAQUA model, which requires a 10% increase in stress values. The eastward (τ_E) and northward (τ_N) components of the wind stress were first used to calculate the total wind stress magnitude (τ):

$$\tau = 1.1\sqrt{\tau_E^2 + \tau_N^2}$$

where the 10% correction factor is applied by multiplying the stress by 1.1. This corrected stress value is then used to determine the friction velocity (u^*) and subsequently the 10-meter wind speed (u_{10}) based on the following steps:

1. ***Friction velocity*** calculation: The friction velocity (u^*) is calculated by:
$$\mathrm{u}^* = \sqrt{\frac{\tau}{\rho_a}}$$
where ρ_a is the density of air, taken as $1.21\ [kg/m^3]$
2. ***Surface roughness*** length calculation: The surface roughness length z_0 is derived using Charnock's relationship:
$$z_0 = \frac{\alpha_M \nu}{u^*} + \frac{\alpha_C u^{*2}}{g}$$
Here, α_M is set to 0.11 (following (Charnock, 1955)), ν is the kinematic viscosity of air $(1.5 * 10^{-5}[m^2/s])$, g is the gravitational acceleration (9.81 $[m/s^2]$), and α_C is the Charnock parameter, set to $0.025\ [-]$.
3. ***Drag coefficient*** calculation: Using z_0, the drag coefficient (C_D) is computed as:
$$C_D = \left(\frac{k}{ln\left(\frac{z}{z_0}\right)}\right)^2$$
where k is the Von Kármán constant (0.4) and z is the reference height (10 $[m]$)
4. ***Pseudo wind speed*** calculation: Finally, the 10-meter wind speed (u_{10}) is derived by:
$$u_{10} = \frac{u^*}{\sqrt{C_D}}$$

This method ensures that the wind speed calculation accurately reflects the conditions specified by the stress fields, with adjustments to match the requirements of the WAQUA model. Additionally, wind direction was computed using the eastward (u_{10}) and northward (v_{10}) wind components from SEAS5, with the direction calculated as:

$$Direction = \left(arctan\left(\frac{-u_{10}}{-v_{10}}\right)\frac{180}{\pi}\right) mod\ 360$$

B.5 Maximum Dissimilarity Algorithm (MDA)

The Maximum Dissimilarity Algorithm (MDA) is a method used to select a representative subset of data from a larger dataset by maximising the dissimilarity between each of the selected points. This technique is particularly useful when the goal is to ensure that the selected points are as diverse as possible, thereby providing a broad representation of the entire dataset (Willett, 1999).

Reasons for applying MDA in this study

MDA is employed in this study as a way to select a set of days that represent the full range of variability present in the dataset, ensuring that the chosen days are spread across the feature space. Unlike random sampling, which may result in highly similar or redundant samples, MDA explicitly focuses on selecting points that are maximally different from each other. This is achieved by calculating pairwise dissimilarities and iteratively choosing the point that maximises the minimum distance from the already selected points.

This method is particularly advantageous in cases where a limited number of samples need to be selected but where it is crucial that these samples reflect the underlying diversity of the full dataset. For instance, in this study, the goal was to select days with extreme surge values from a large dataset of surge predictions, ensuring that these selected days cover a wide range of surge scenarios.

MDA - mathematically explained

The MDA process follows these steps:

1. ***Initial selection*** - The first point is chosen from the dataset. This can be done either randomly or based on specific criteria, such as selecting a day with the highest surge.
2. ***Iterative selection*** - For each subsequent point, the algorithm computes the dissimilarity (distance) between the candidate point and the already selected points. The goal is to select the point that is maximally dissimilar from the set of previously chosen points. Mathematically, this is done by selecting the point x_{next} that maximises the minimum distance to all previously selected points:
$$x_{next} = \arg max_{x_i \in \{x_i, \ldots, x_k\}} min_{j=i}^{k} \, d(x_i, x_j)$$
where:
   - X – is the full dataset of points
   - $\{x_1, x_2, \ldots, x_k\}$ – are the points already selected
   - $d(x_i, x_j)$ – is the dissimilarity (typically the Euclidean distance) between points x_i and x_j
3. ***Termination*** - Repeat this process until the desired number of points have been selected.

MDA - visually explained

To illustrate how the Maximum Dissimilarity Algorithm (MDA) works, consider a simplified 2D dataset of points, as shown in Figure B-8. The goal of MDA is to iteratively select data points such that each selected point is as different as possible from the ones already chosen. The algorithm begins by selecting an initial point (either randomly or preselected), after which it iterates through the remaining points, choosing the one that is farthest from the previously selected points.

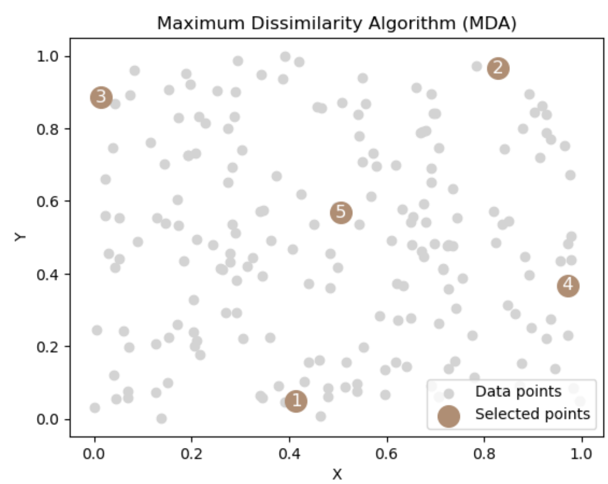


Figure B-8: Visual representation of the Maximum Dissimilarity Algorithm (MDA) applied to a 2D dataset. The points labeled 1 to 5 represent the order of selection based on maximising the minimum distance from the already selected points.

In this simplified example the first point is randomly selected from the dataset, and labelled 1 in the plot. This point is used as the starting reference for subsequent selections. The second point is selected to be the one that is most dissimilar to the first point. This is calculated by finding the point that has the maximum minimum distance to the already selected points (in this case, just the first point). The second point is labelled 2. The algorithm then continues iterating through the dataset, always selecting the point that has the largest minimum distance to the set of previously selected points.

The key takeaway from this visualisation is that MDA ensures that the selected points are spread across the feature space, thus maximising the diversity of the chosen subset. This method is particularly useful when one wants to cover as much of the data's variability as possible with a smaller sample.

B.6 Stratified sampling

Stratified sampling is a method that divides a dataset into subgroups, or "strata," that share specific characteristics. The aim is to ensure that each subgroup is adequately represented in the sample (Hayes, James, & Beer, 2024).

Reasons for applying stratified sampling in this study

Stratified sampling is employed in this study to ensure that extreme surge events are adequately represented in the final sample. Extreme surge categories are rare, and without stratified sampling, random sampling might result in these conditions being underrepresented.

By dividing the dataset into nine surge categories (ranging from very low to extreme), stratified sampling guarantees that each category is represented equally. This approach allows the study to capture the full variability of surge conditions, with particular emphasis on ensuring extreme surges are included. The method ensures that the selected days reflect the range of surge scenarios, which might otherwise be overlooked in random sampling.

Stratified sampling explained

Stratified sampling involves several steps to ensure that each subgroup (or stratum) is adequately represented in the sample. The process can be broken down as follows:

1. **Divide the data into strata -** The dataset is first divided into different strata based on a specific variable, such as surge values in this study. In the case of surge data, the strata are created by defining bins (ranges) of surge values. For example, bins may be defined for surge values less than -2 meters, between -2 and -1 meters, and so on.
2. **Determine the number of samples to be selected from each stratum -** In this study, the number of samples selected from each stratum is the same for all strata, regardless of how many data points are in each surge category. This ensures that extreme surge categories are adequately represented. For instance, if 100,000 days are to be selected, an equal number of samples (in this case, 11,111 days) is chosen from each of the 9 defined surge bins.
3. **Random sampling within each stratum -** Once the strata have been defined and the sample size for each stratum has been determined, the algorithm selects a random sample from each stratum. This ensures that the selected points are representative of the variations within each surge level. Mathematically, this process can be represented as:
$$Sample_i = Randomly\ select\ n_i\ from\ stratum\ i$$
where n_i represents the number of samples selected from stratum i.

B.7 Entropy and Empirical Cumulative Distribution Functions (ECDFs)

Shannon entropy is a fundamental concept in information theory that quantifies the amount of uncertainty or variability in a probability distribution (Shannon, 1948). It measures how unpredictable or dispersed a dataset is. In this study, Shannon entropy is used to assess the variability of storm surge distributions associated with different weather patterns. A higher entropy value indicates a more diverse range of storm surge values within a given pattern, while a lower entropy suggests that surge values are more concentrated around certain values.

Mathematical definition

For a discrete probability distribution $P = \{p_1, p_2, \ldots, p_n\}$, the Shannon entropy is given by (Shannon, 1948):

$$H(P) = -\sum_i p_i \log_2 p_i$$

where:

- $H(P)$ – is the Shannon entropy
- p_i – represents the probability of each discrete value x_i occurring in the dataset.

Entropy is highest when the probabilities are uniformly distributed (i.e., all outcomes are equally likely) and lowest when the probabilities are highly concentrated around specific values. This concept is illustrated in Figure B-9, where the left panel shows a uniform distribution with the highest possible entropy, the middle panel depicts a moderate entropy distribution following a normal shape, and the right panel demonstrates a highly concentrated distribution with the lowest entropy.

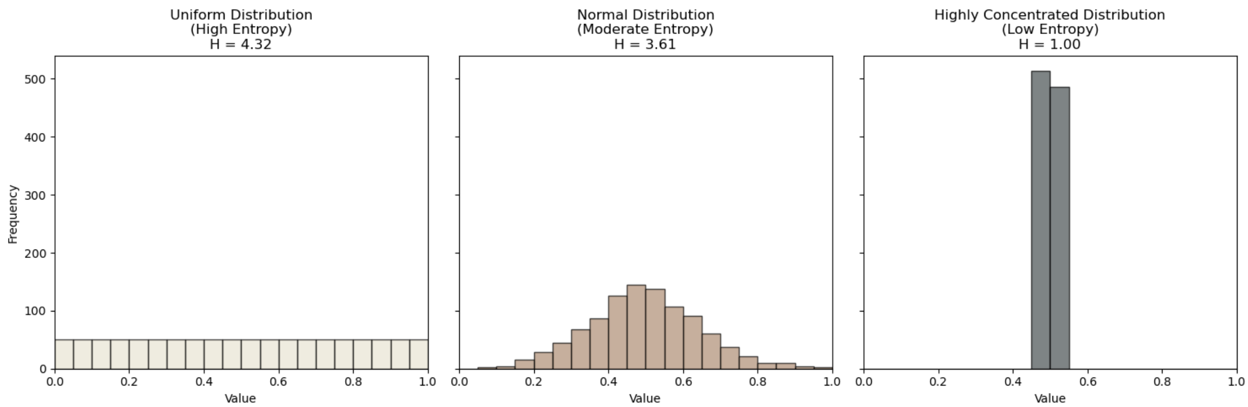


Figure B-9: Example of entropy in different probability distributions.

Empirical Cumulative Distribution Functions (ECDFs)

To compute entropy, it is necessary to estimate the probability distribution of storm surge values for each weather pattern. In this study, this is done using Empirical Cumulative Distribution Functions (ECDFs).

The ECDF of a dataset provides an estimate of the cumulative probability of a random variable by calculating the fraction of observations that fall below a given value. Mathematically, for a sample of n observations $x_1, x_2, \dots, x_n$, the ECDF is defined as (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005):

$$F(x) = \frac{1}{N} \sum_{i=1}^{n} 1(x_i \leq x)$$

where:

- $F(x)$ – is the cumulative probability up to x
- $1(x_i \leq x)$ – is an indicator function that equals 1 if $x_i \leq x$ and 0 otherwise
- n – is the total number of observations

The ECDF is a non-parametric estimator of the cumulative distribution function (CDF), meaning it does not assume a specific underlying probability distribution.

A visual example of the construction of an ECDF is illustrated in Figure B-10. The left panel shows a dummy dataset as a histogram, displaying the frequency of fictious storm surge values. The right panel presents the corresponding ECDF, which accumulates the proportion of observations below each threshold.

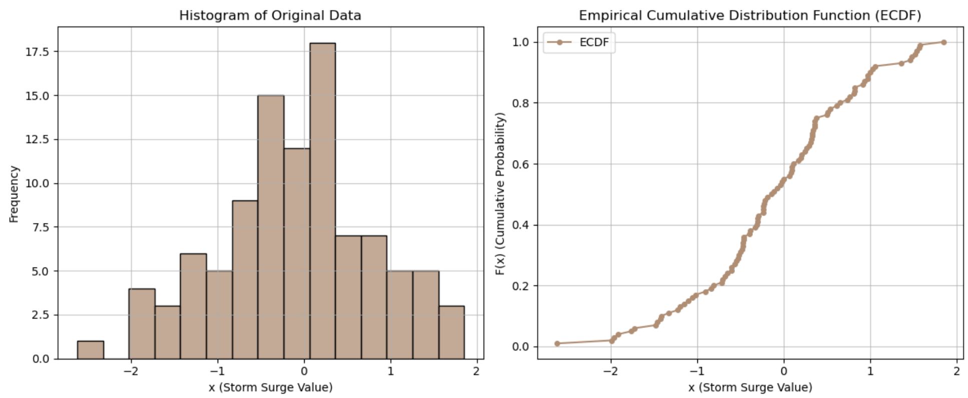


Figure B-10: Visual example of ECDF. Histogram of the data (left) and the corresponding ECDF (right).

Applications in this study
In this study, Shannon entropy is computed for each Najda weather pattern based on the probability distribution of storm surge values. The probability distributions are obtained using Empirical Cumulative Distribution Functions (ECDFs), which estimate the probabilities of different surge values occurring within each weather pattern.

To compute the entropy:

1. Storm surge values are sorted within each Najda weather pattern.
2. The ECDF is computed to determine the proportion of data points below each surge value.
3. Probabilities are extracted from the ECDF.
4. Shannon entropy is calculated using the formula above.

By computing Shannon entropy for each pattern, this study identifies which weather patterns are associated with the most diverse and uncertain surge outcomes.

Interpretation of results
The entropy value may be interpreted as follows:

- High entropy patterns exhibit greater variability in storm surge values, meaning that the storm surges associated with those weather patterns are less predictable and more widely spread.
- Low entropy patterns have more concentrated storm surge values, meaning that the storm surge levels for those patterns are more consistent and predictable.

B.8 Kullback-Leibler (KL) divergence

The Kullback-Leibler (KL) divergence is a fundamental concept in information theory that measures the difference between two probability distributions (Bishop, 2006). Kl divergence is also known as relative entropy (Braverman, 2011). In this study, KL divergence is used to quantify how different the storm surge distributions are between pairs of Najda weather patterns. A low KL divergence between two patterns indicates that their storm surge distributions are similar, whereas a high KL divergence suggests substantial differences in their surge behaviour.

Mathematical definition
For two probability distributions, P (true distribution) and Q (approximate distribution), the KL divergence is defined as (Bishop, 2006):

$$D_{KL}(P||Q) = \sum_i P(i) \log_2 \frac{P(i)}{Q(i)}$$

where:

- $D_{KL}(P||Q)$ – represents the divergence from Q to P
- $P(i)$ – is the probability of event i in the true distribution
- $Q(i)$ – is the probability of event i in the comparison distribution

KL divergence is asymmetric, meaning that $D_{KL}(P||Q) \neq D_{KL}(Q||P)$. This means the measure is not a true "distance" metric but rather an indication of how much information is lost when approximating P using Q.

Computing KL divergence

In this study, KL divergence is applied to compare the storm surge probability distributions associated with different Najda weather patterns. The probability distributions of storm surges are estimated using Kernel Density Estimation (KDE), a non-parametric technique that smooths the observed surge values to approximate a continuous probability density function (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005). This process is illustrated in Figure B-11, where a histogram of storm surge values is overlaid with the corresponding KDE curve. The KDE provides a smooth estimate of the probability density, avoiding the binning artifacts of histograms while capturing the underlying distribution more accurately.

The steps involved in computing KL divergence are:

1. Define a common grid of surge values spanning the full range of observations.
2. Estimate probabilities for each weather pattern by calculating surge occurrence frequencies over this grid.
3. Compute KL divergence between each pair of patterns using the formula above.
4. Construct a KL divergence matrix, where each entry represents the divergence between two patterns.

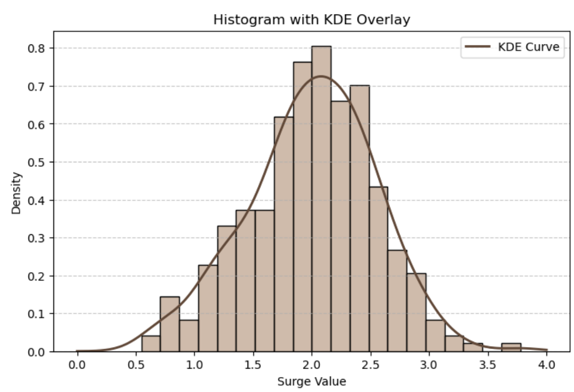


Figure B-11: Visual example of KDE curve. The histogram shows discrete surge value counts, while the KDE curve provides a smooth probability density estimate.

To facilitate interpretation, the computed KL divergence values are visualised as heatmaps in the main report, where darker colours represent stronger distinctions between weather patterns.

Handling of infinite KL divergence

In cases where one weather pattern completely lacks certain surge values that appear in another pattern, KL divergence becomes mathematically infinite. To ensure meaningful visualisation and comparison, infinite values are replaced with a large finite value for plotting purposes. This allows the heatmaps to represent extreme differences while avoiding issues with numerical instability**.**

Interpreting KL divergence in this study

KL divergence may be interpreted as follows:

- ***Low*** KL divergence between two weather patterns means that their storm surge distributions are similar, suggesting that those patterns produce comparable surge behaviour.
- ***High*** KL divergence indicates that the storm surge distributions are significantly different, meaning that one weather pattern is associated with a very different range of surge values than another.

- ***Infinite*** KL divergence occurs when one distribution assigns zero probability to a surge value that is present in the other distribution. This signifies complete dissimilarity between the patterns.

These three scenarios are illustrated in Figure B-12. The left plot shows two similar distributions with minor differences, resulting in low KL divergence. The middle plot represents high KL divergence, where the distributions differ significantly but still share some overlap. The right plot illustrates infinite KL divergence, where the two distributions have completely non-overlapping support, leading to an undefined or infinite KL value.

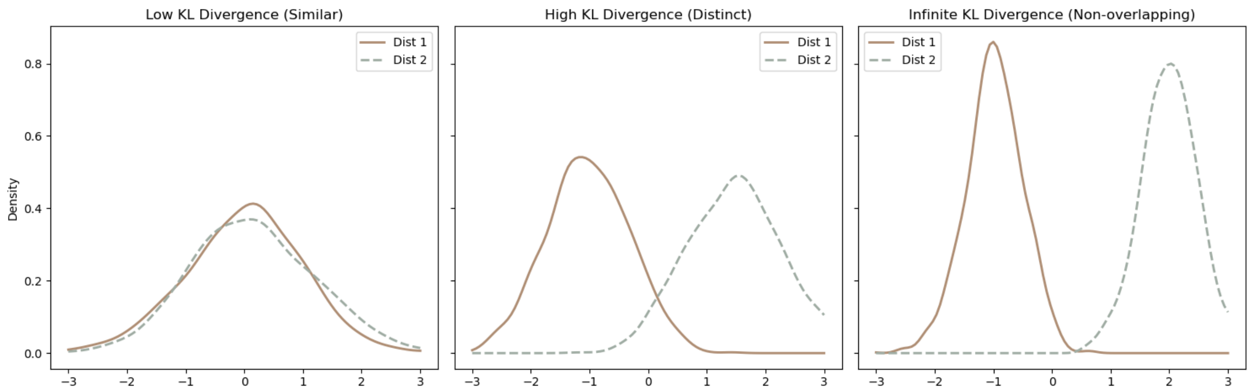


Figure B-12: Examples of KL divergence scenarios. Left: Low KL divergence—two similar distributions. Middle: High KL divergence—distinct but overlapping distributions. Right: Infinite KL divergence—completely non-overlapping distributions.

By analysing the KL divergence matrix, the redundancy and distinctiveness of the weather patterns may be assessed in terms of their associated storm surge distributions. Patterns with very low KL divergence may be considered redundant, while patterns with high KL divergence indicate unique storm surge behaviour.

Appendix C: Surge return period analysis

This appendix briefly describes a small internal study conducted by WMCN to link return periods to surge values along the Dutch coast. The analysis, shared generously by WMCN, provides valuable insights into the frequency of storm surges at several coastal locations.

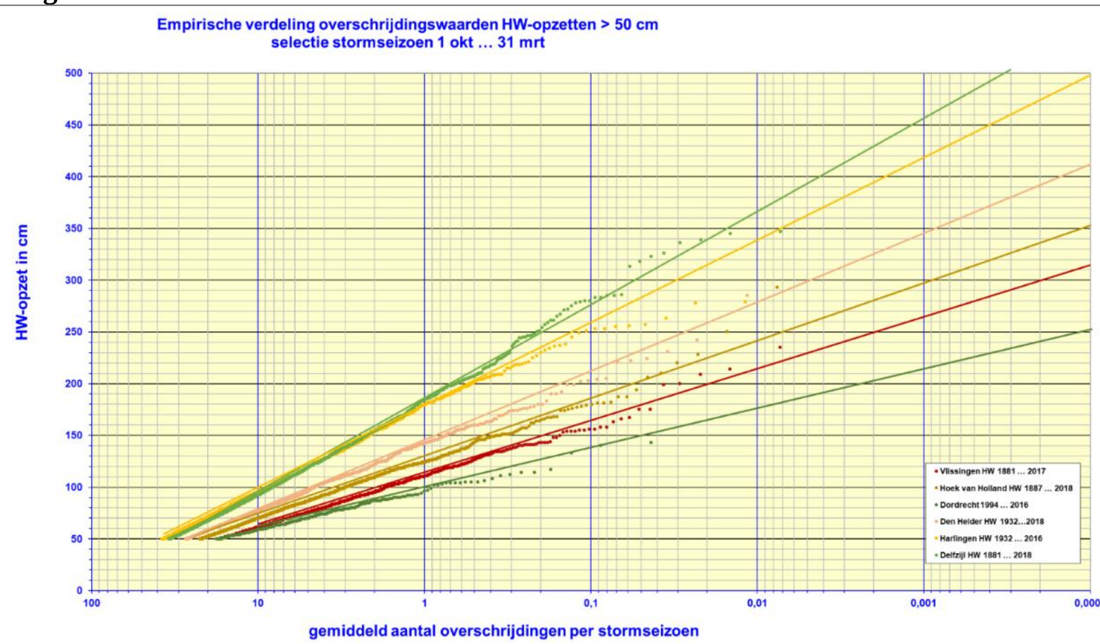
The study focused on the storm season, defined as the period from 1st October to 31st March, and included only storm surges exceeding 50 cm. For each location, the return period (x-axis) was linked to the high-water surge values (y-axis). The results are visualised in Figure C-0-1 figure below, where each colour represents a different location, and the legend indicates the length of the historical record used for the analysis. For example, at Vlissingen, records spanning from 1881 to 2017 were utilised.

This analysis highlights the variability in return periods for different surge thresholds across locations, reflecting both regional differences and the influence of long-term data availability. Based on these results, Table C-0-1 summarises the approximate surge values for two key metrics:

- Surge threshold = 1.5 m: This threshold was used in the Method A of Phase II to capture significant but not overly rare storm surge events.
- Return period = 1 year: While not directly used in the methodological framework of Phase II, incorporating return period-based surge values into future forecasting models is recommended. This is discussed further in Section 12.2.1, where improvements to weather pattern-based forecasting are proposed.

Table C-0-1: Approximate return period for various surge thresholds across multiple locations along the Dutch coast

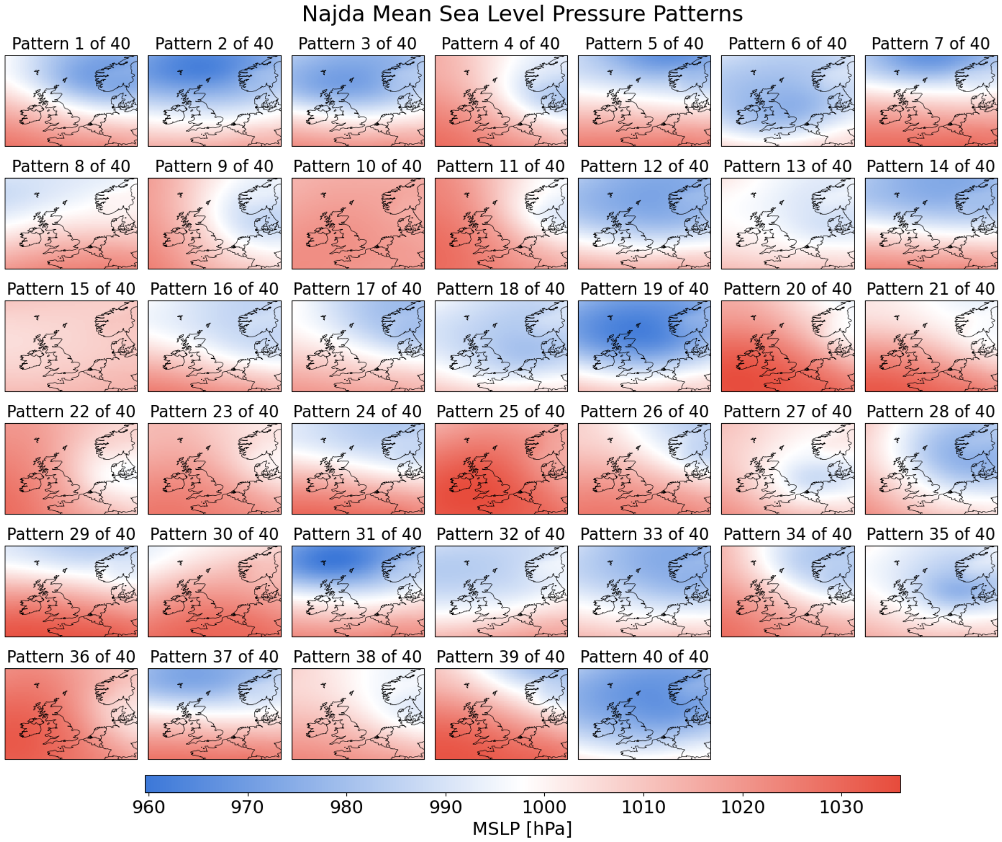
Location	Return Period [year] for surge = 1.5 [m]	Surge threshold [cm] for Return Period = 1 [year]
Delfzijl	2.30	185
Harlingen	2.00	180
Den Helder	0.80	142
IJmuiden	0.50	130
Hoek van Holland	0.40	125
Vlissingen	0.17	110

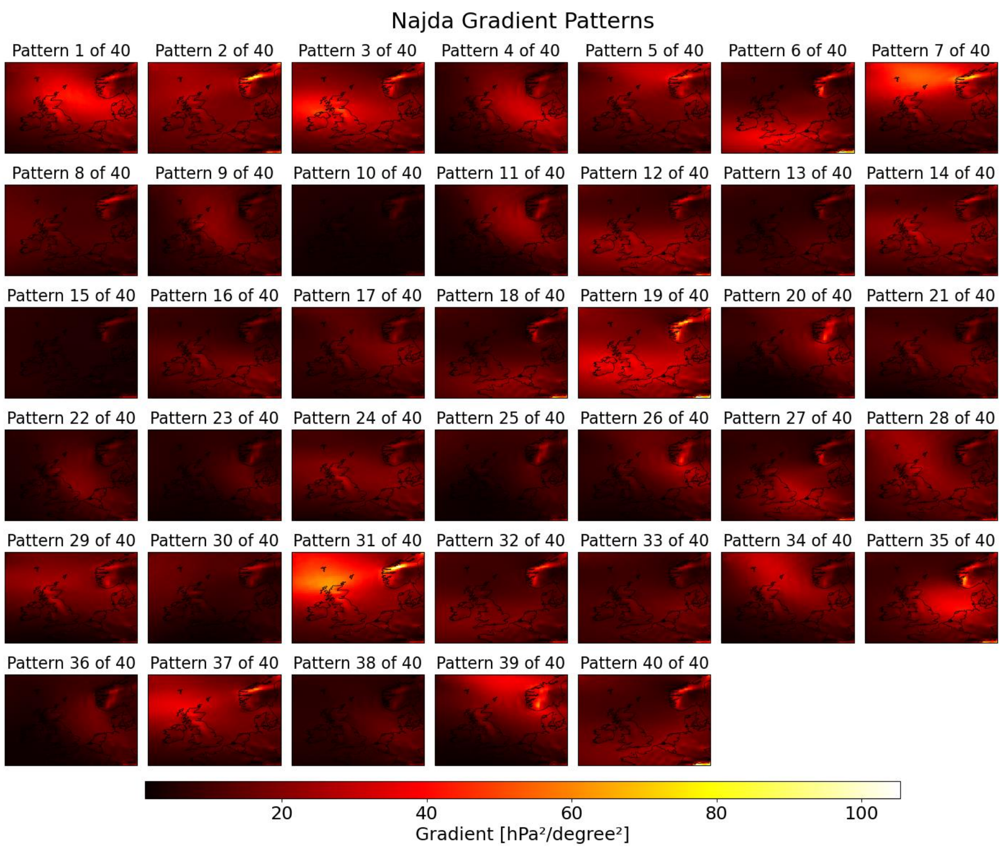


Appendix D: Generated weather patterns

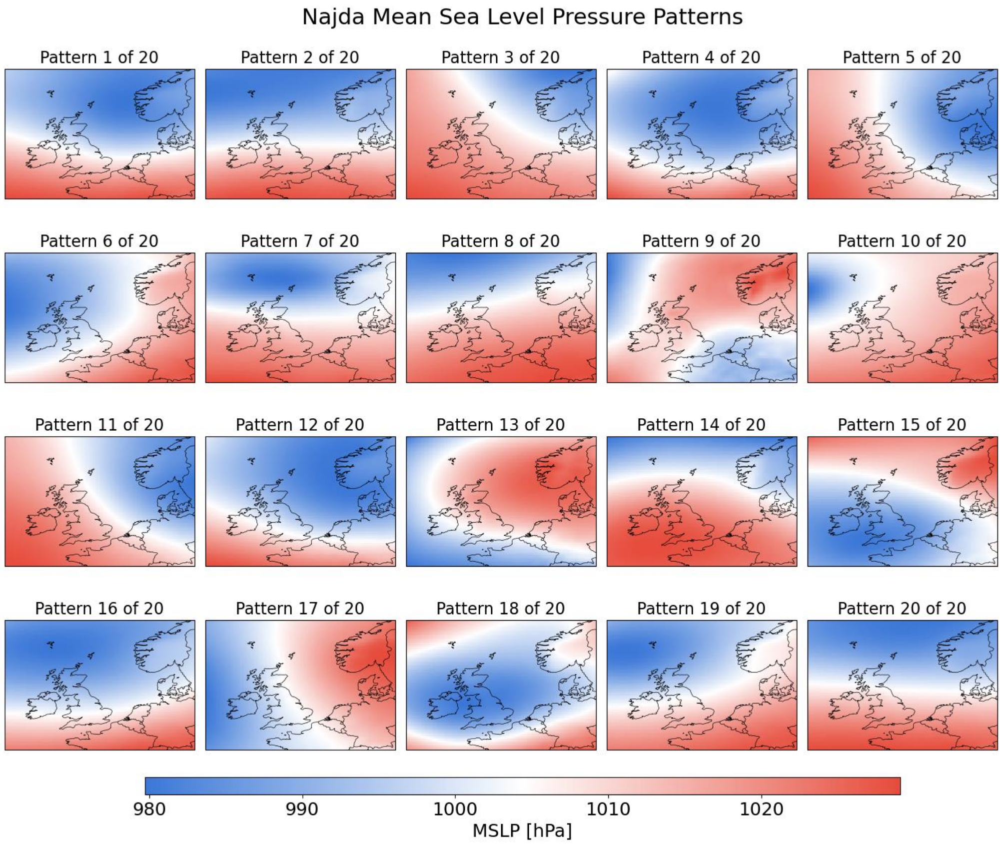
This appendix provides larger versions of the mean sea level pressure (MSLP) and pressure gradient weather pattern plots for Methods A, B, and C, allowing for a more detailed examination of the generated patterns beyond the smaller visualisations presented in the main text.

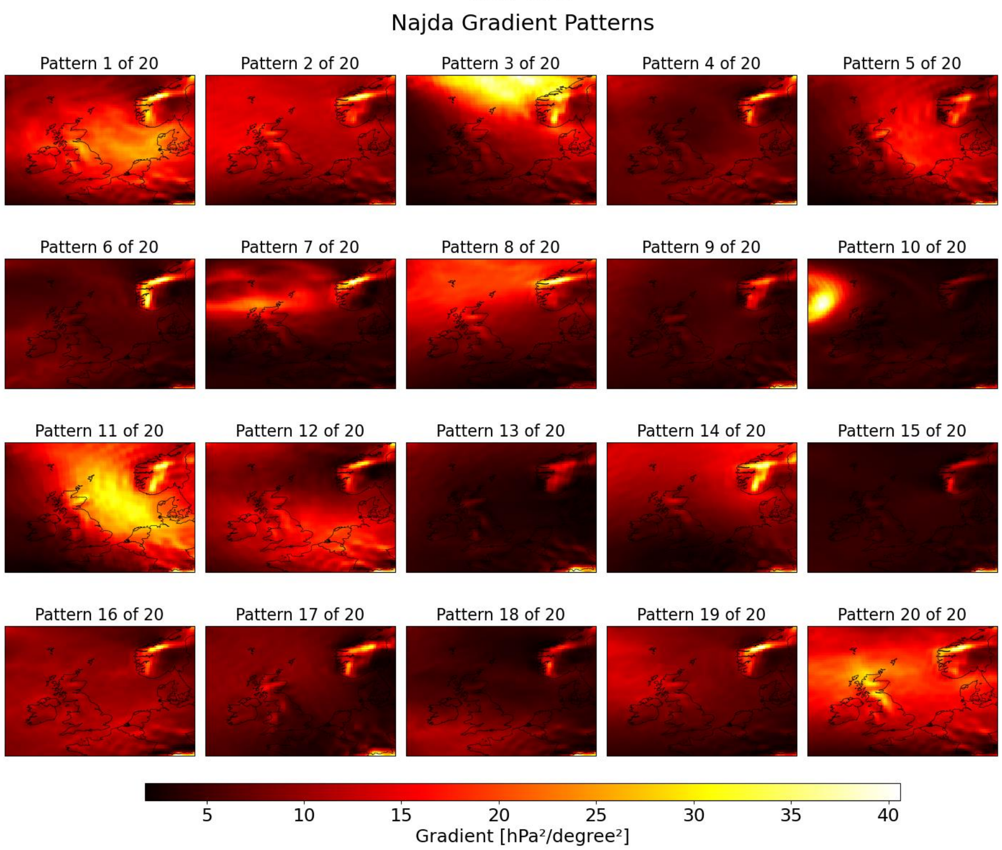
Method A: Surge threshold



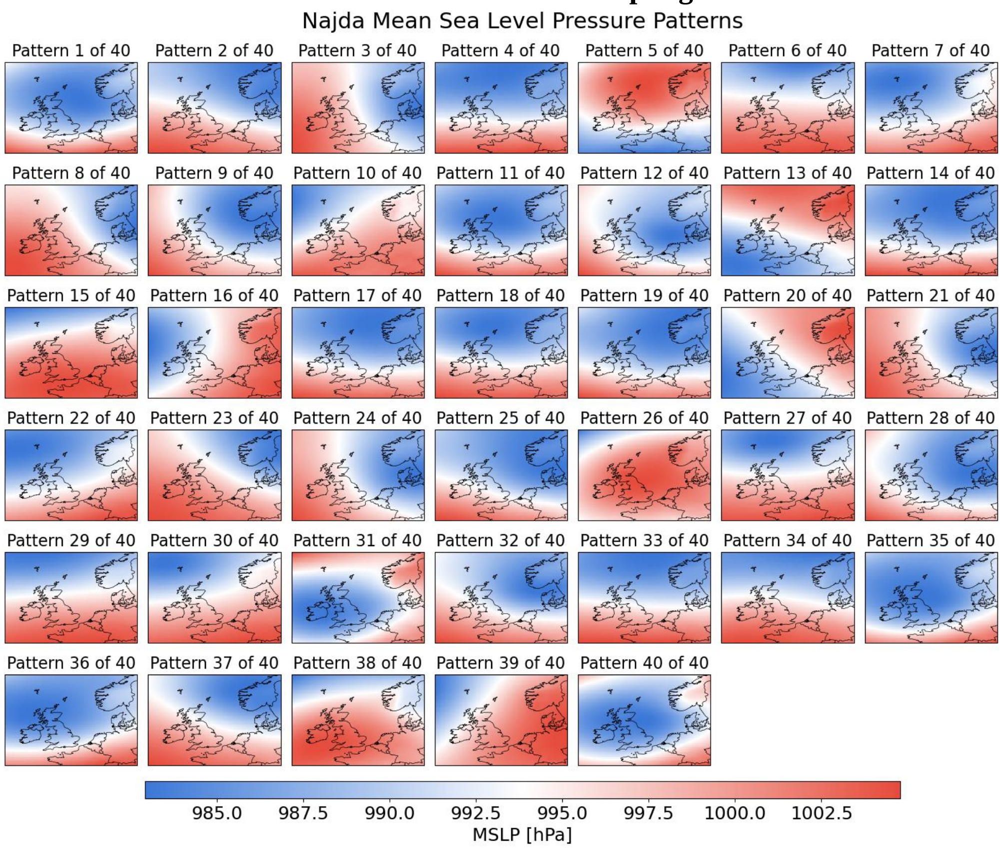


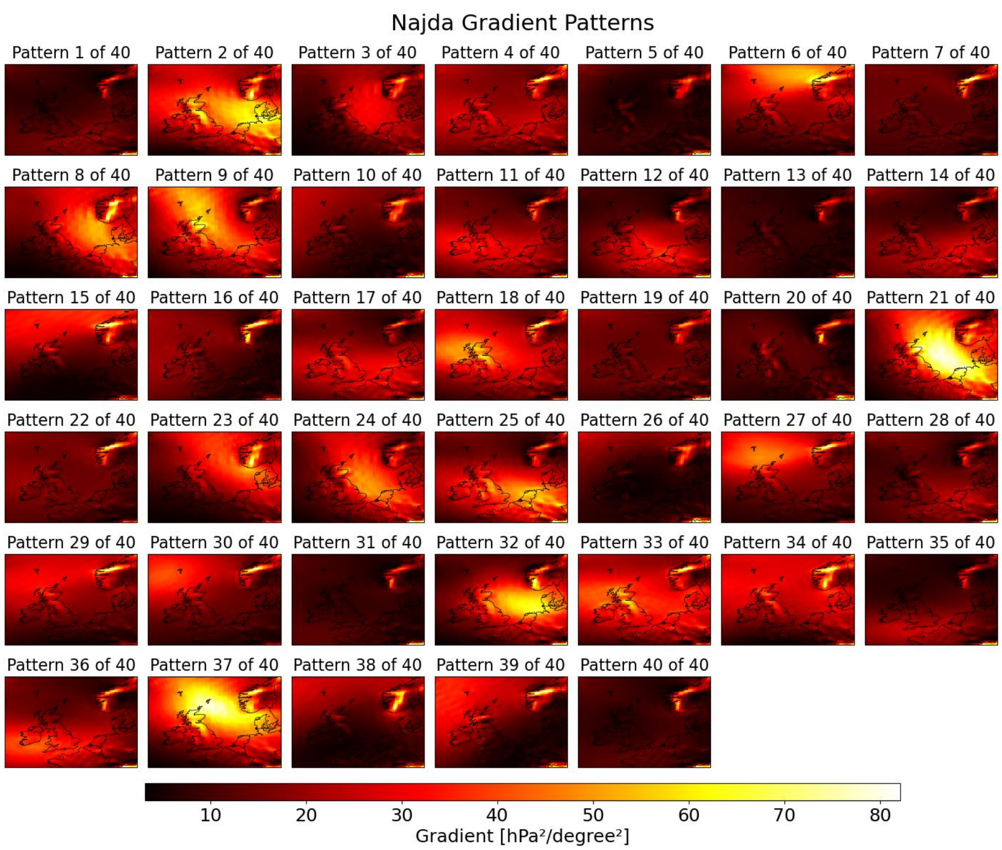
Method B: MDA


Najda Mean Sea Level Pressure Patterns
Pattern 1 of 20
Pattern 2 of 20
Pattern 3 of 20
Pattern 4 of 20
Pattern 5 of 20
Pattern 6 of 20
Pattern 7 of 20
Pattern 8 of 20
Pattern 9 of 20
Pattern 10 of 20
Pattern 11 of 20
Pattern 12 of 20
Pattern 13 of 20
Pattern 14 of 20
Pattern 15 of 20
Pattern 16 of 20
Pattern 17 of 20
Pattern 18 of 20
Pattern 19 of 20
Pattern 20 of 20
980
990
1000
1010
1020
MSLP [hPa]


Najda Gradient Patterns
Pattern 1 of 20
Pattern 2 of 20
Pattern 3 of 20
Pattern 4 of 20
Pattern 5 of 20
Pattern 6 of 20
Pattern 7 of 20
Pattern 8 of 20
Pattern 9 of 20
Pattern 10 of 20
Pattern 11 of 20
Pattern 12 of 20
Pattern 13 of 20
Pattern 14 of 20
Pattern 15 of 20
Pattern 16 of 20
Pattern 17 of 20
Pattern 18 of 20
Pattern 19 of 20
Pattern 20 of 20
5
10
15
20
25
30
35
40
Gradient [hPa²/degree²]

Method C: Stratified sampling





Appendix E: Dendrograms

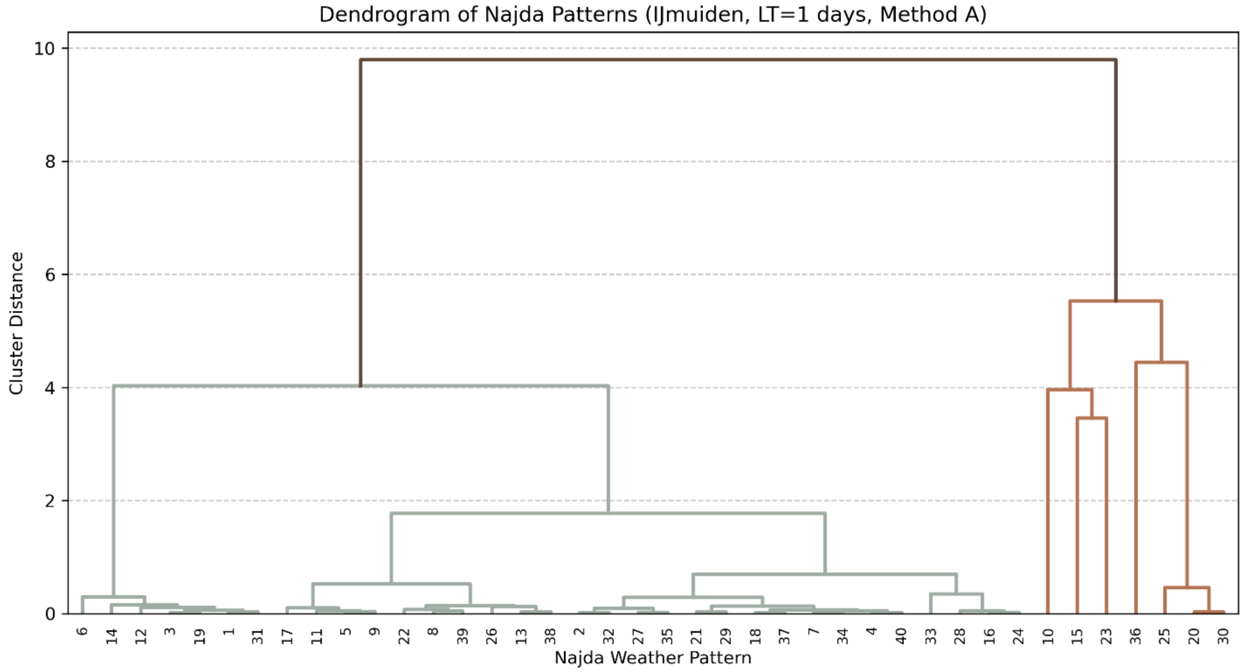
This appendix presents larger versions of the dendrograms for each tried method (A, B and C), providing a clearer view of the hierarchical clustering structure.

A dendrogram is a tree-like diagram used to visualise the arrangement of clusters in hierarchical clustering. It shows how individual elements (in this case, Najda weather patterns) are progressively merged into larger clusters based on their similarity.

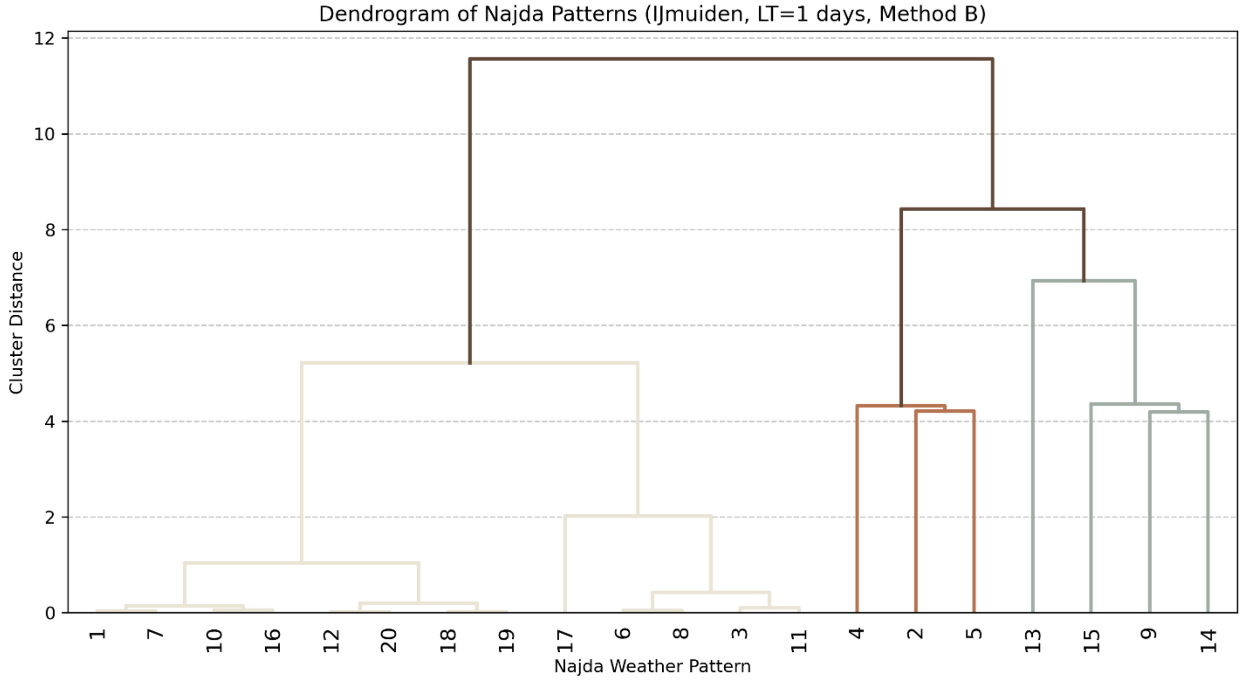
The vertical axis represents the cluster distance, which quantifies the dissimilarity between merged clusters. The higher the merging occurs on the dendrogram, the more dissimilar the clusters are. The horizontal axis represents the individual weather patterns, which are gradually grouped together.

In this study, the dendrogram helps to assess how distinct or redundant the generated weather patterns are. If patterns merge at very low distances, it suggests redundancy, meaning some clusters may be too similar. Conversely, if patterns remain separate until high distances, it indicates greater distinctiveness among the clusters.

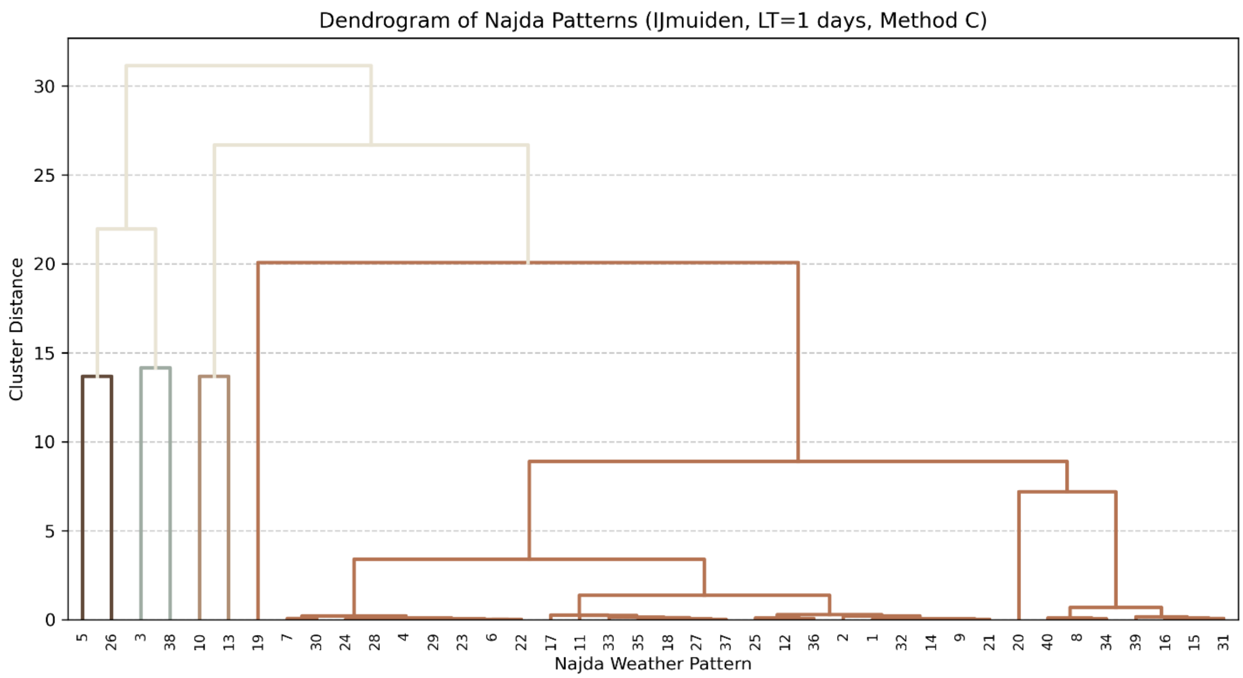
Method A: Surge threshold



Method B: MDA



Method C: Stratified sampling



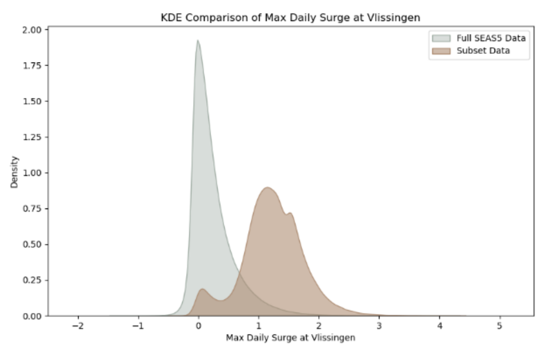
Appendix F: KDE plots

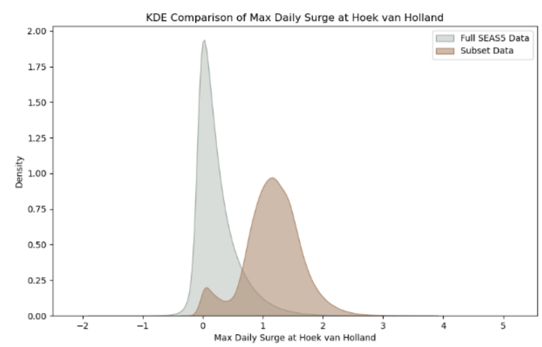
This appendix provides the Kernel Density Estimation (KDE) plots for all six locations considered in the study. While the main report focuses on the results for IJmuiden, as methods B and C are applied based on surge data from this location, the following KDE plots present the distribution of maximum daily surge values for all six locations under each of the three data selection methods.

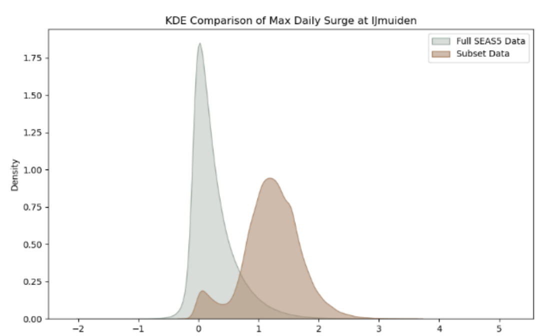
- Method A (Surge threshold) focuses on selecting days with high surges, emphasising extreme weather conditions.
- Methods B and C (MDA and Stratified sampling) aim to capture a broad surge distribution, including both high and low surge days, ensuring a more balanced representation across surge categories.

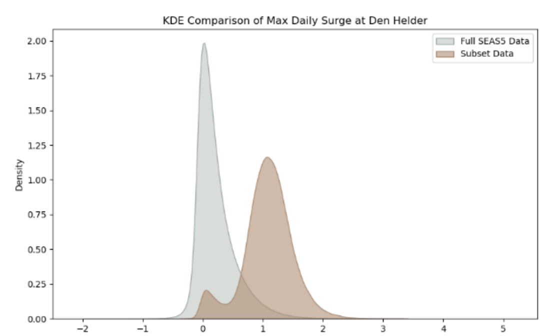
These plots visually compare the surge distribution of the selected subset data against the full SEAS5 dataset, showcasing how the data selection methods influence the resulting subset and their representation of various surge conditions.

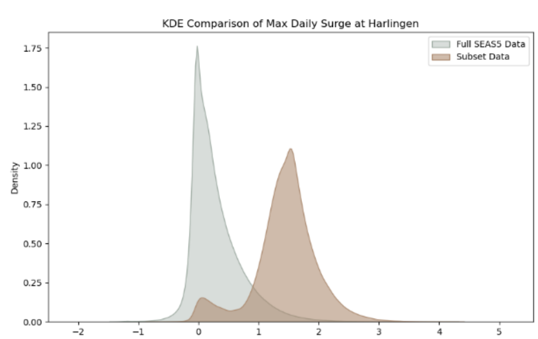
Method A: Surge threshold

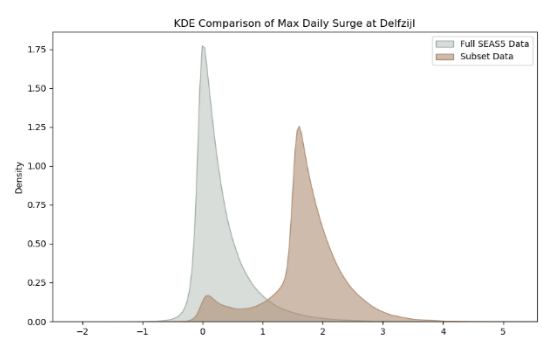








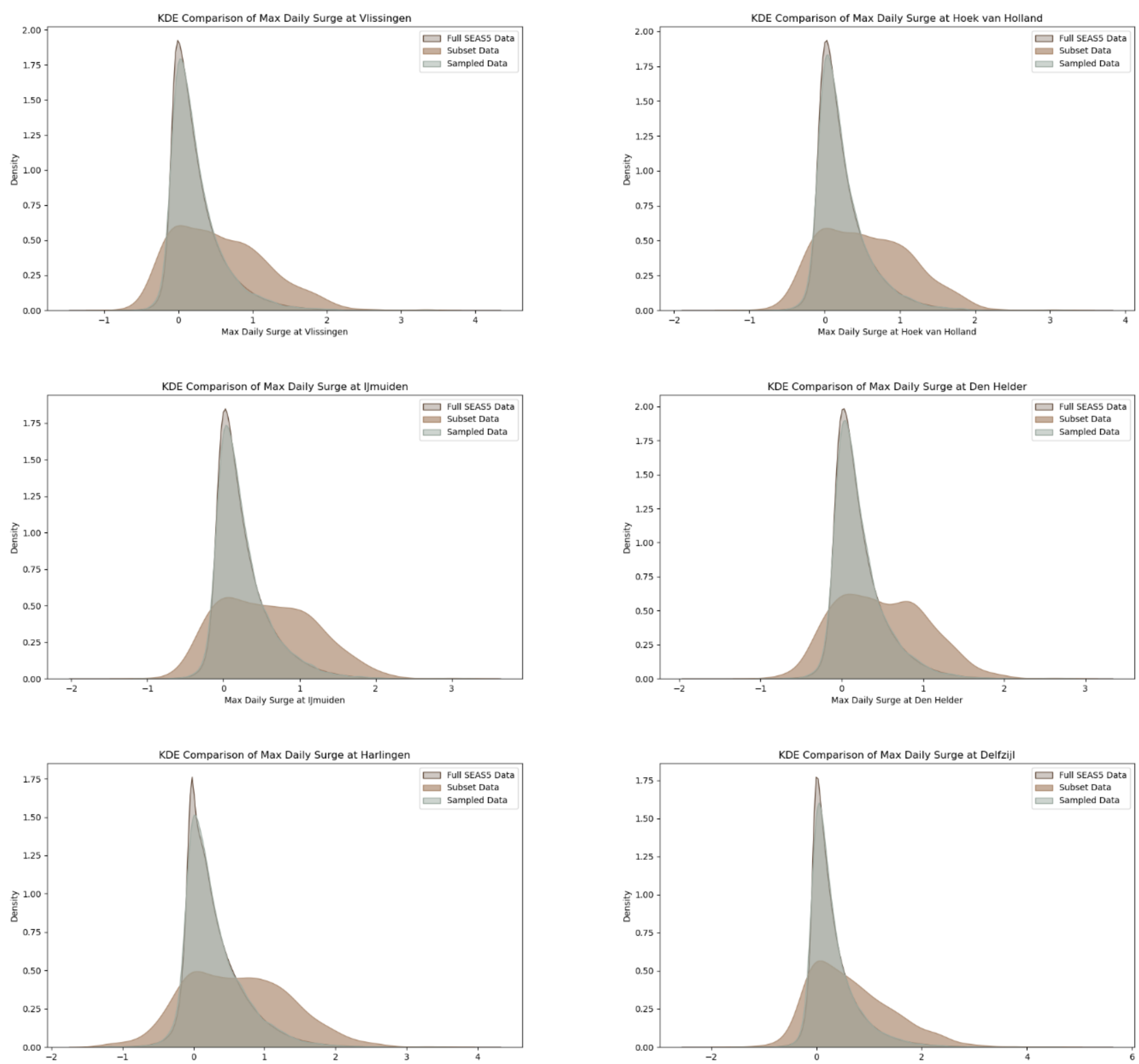




Based on the KDE plots for Method A, a few observations/conclusions can be made:

i. The fixed surge threshold approach prioritises the inclusion of high surge days, making sure extreme surge events are adequately represented in the dataset.
ii. The surge in Delfzijl is inherently higher than in other locations like Vlissingen. As a result, many days were selected based on the surge exceeding the 1.5-meter threshold in Delfzijl, while on the same days, surge values in Vlissingen may have been much lower, not exceeding the threshold. This could skew the selection process toward locations with higher surge levels.

Method B: MDA

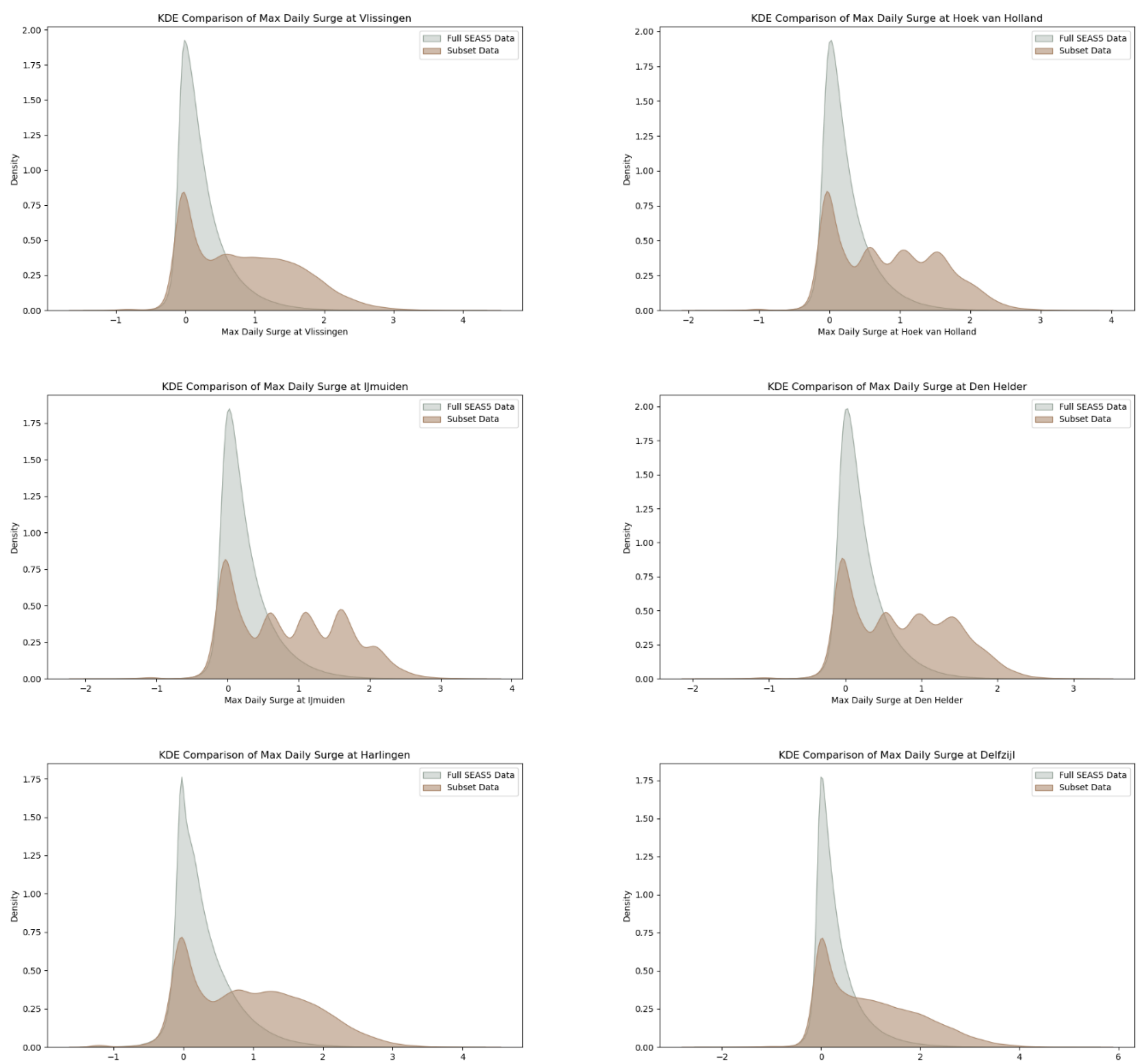


Based on the KDE plots for Method B, a few observations/conclusions can be made:

i. Balanced surge distribution: Method B attempts to capture a broader range of surge values by maximising the dissimilarity between selected days. The KDE plots show that, compared to Method A, the sampled data in this approach better reflects a balanced distribution of surge values, including both high and low surges, for all locations.

ii. Focus on extreme events maintained: While MDA promotes diversity in the selection process, it still ensures that extreme surge days are adequately represented. This is evident in locations like IJmuiden, where extreme surge values are included in the sampled data without overwhelming the lower surge values. As IJmuiden was used for the MDA selection, this result was expected for that location. However, it is notable that for the other locations, the surge distribution is also quite broad, despite the data not being selected based on these locations. This suggests that selecting based on a single location may still provide a broad and representative distribution of surge values for other locations along the Dutch coast as well.

Method C: Stratified sampling



Based on the KDE plots for Method C, a few observations/conclusions can be made:

i. Concentrated surge distribution for IJmuiden: Method C results in a KDE plot for IJmuiden that shows multiple peaks, indicating that the stratified sampling method captures a range of surge values across different surge categories (e.g., "low," "medium," "high," etc.). While there are peaks corresponding to both lower and higher surge values, the distribution is not uniform within each surge category. This is likely due to an overrepresentation of certain surge values within each category.

ii. Peaked distributions for nearby locations: The KDE plots for Hoek van Holland and Den Helder also exhibit relatively peaked distributions, similar to IJmuiden. This can be attributed to the fact that these locations are relatively close to IJmuiden, and therefore, their surge behaviour is likely to be similar. As the stratified sampling is based on surge data from IJmuiden, the resulting distributions for these locations show a similar structure.

iii. Smoother surge distributions for more distant locations: For Delfzijl and Vlissingen, which are farther from IJmuiden, the KDE plots appear smoother. This could be due to differences in the surge behaviour at these locations compared to IJmuiden. As the surge events at these more distant locations may not exhibit the same concentration of surge values within certain bins, the stratified sampling captures a more even distribution of surge values across the categories, resulting in smoother plots without the pronounced peaks observed in IJmuiden and the nearby locations.

Appendix G: Boxplots

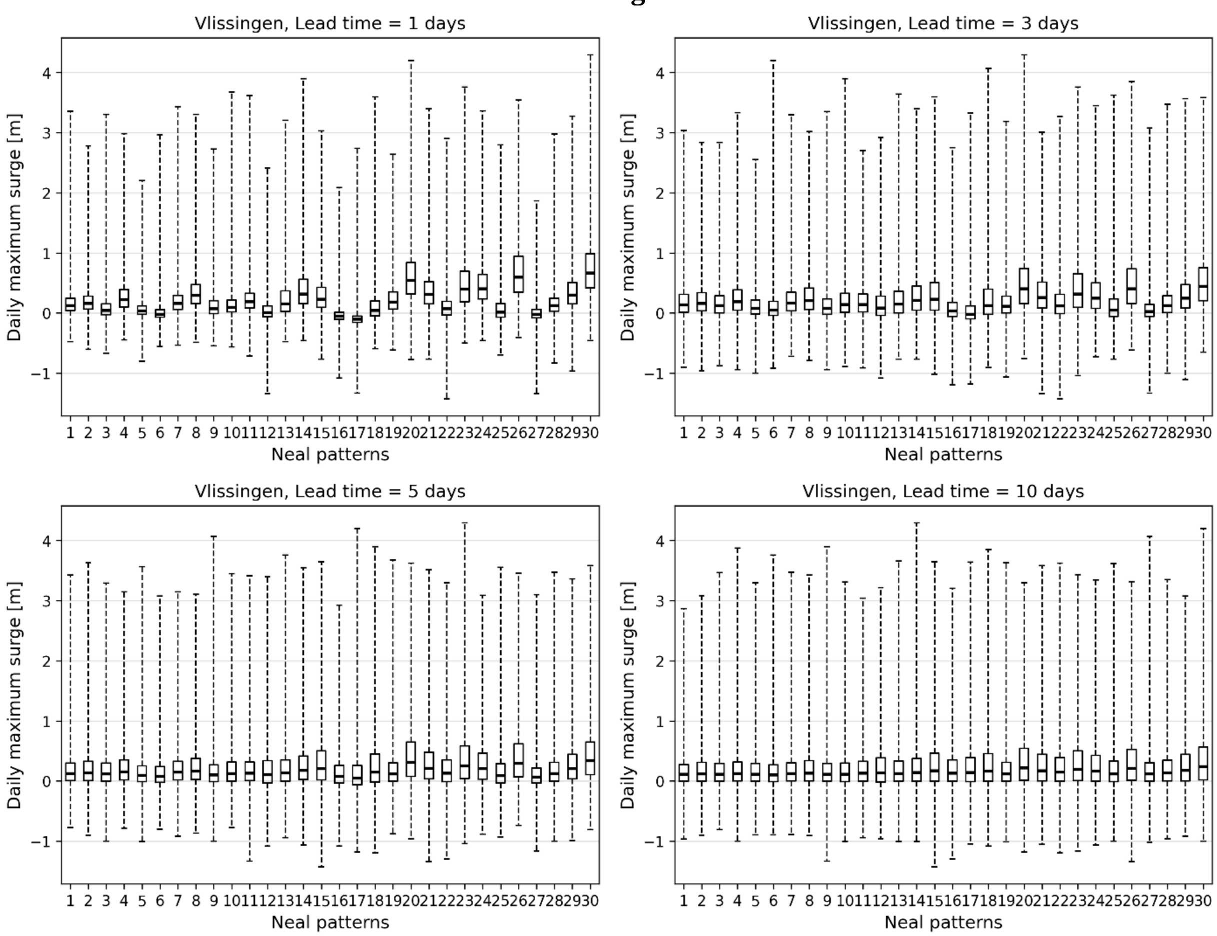
This appendix presents the boxplots illustrating the distribution of daily maximum surge values for different weather patterns, for both Neal patterns and the custom-generated patterns (Methods A, B, and C). These plots offer a comparative overview of surge behaviour across various locations along the Dutch coast and at different lead times.

While the main report focuses on the surge distribution for IJmuiden, the analysis here showcases the results for the other considered locations: Vlissingen, Hoek van Holland, Den Helder, Harlingen and Delfzijl.

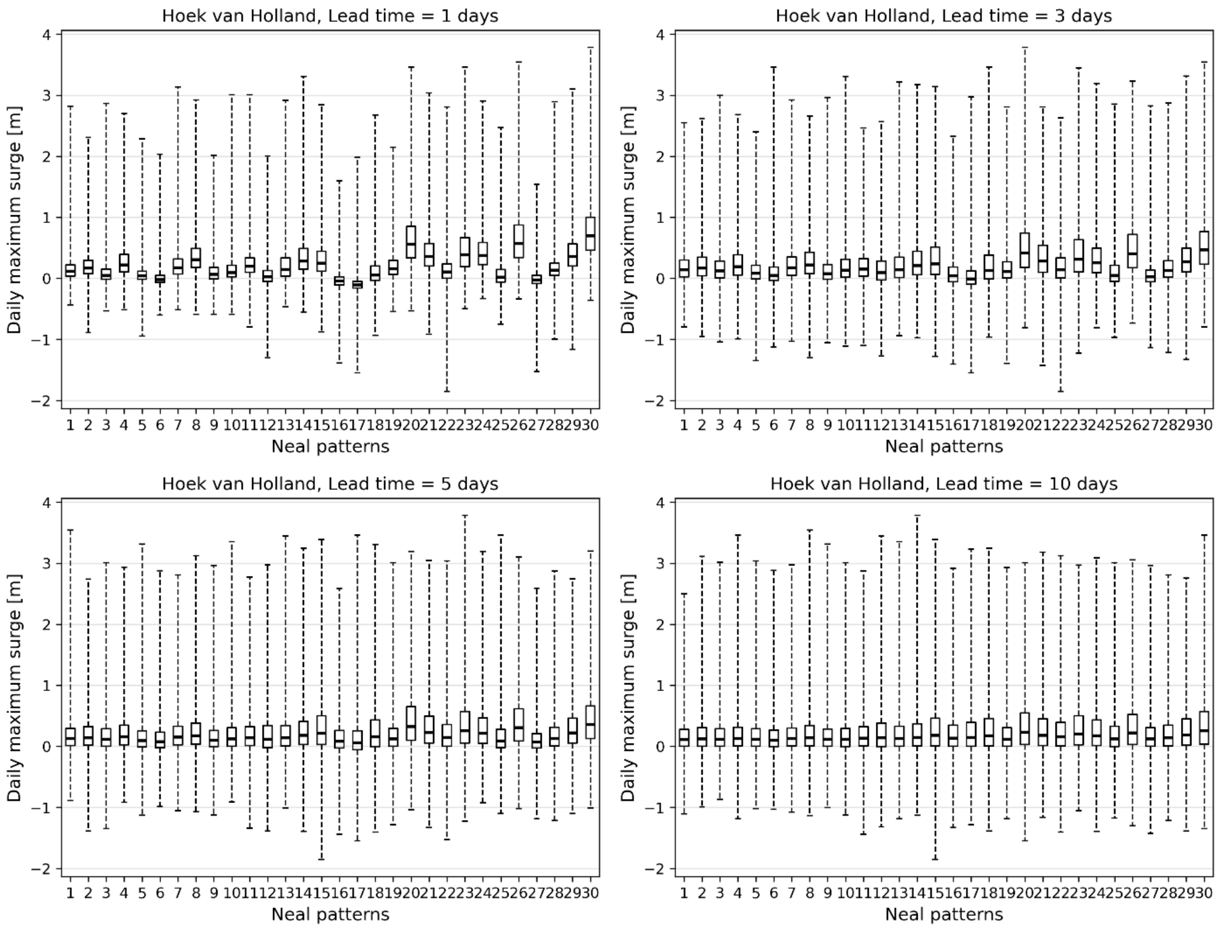
The analysis is presented for lead times up to 10 days. The 15-day lead time boxplot has been excluded from this appendix, as it has been observed that, from a lead time of 10 days onward, the surge distributions across the weather patterns become increasingly similar. This indicates that the predictive usefulness of the surge data at longer lead times is limited, as the influence of the weather patterns converges over time.

Neal patterns

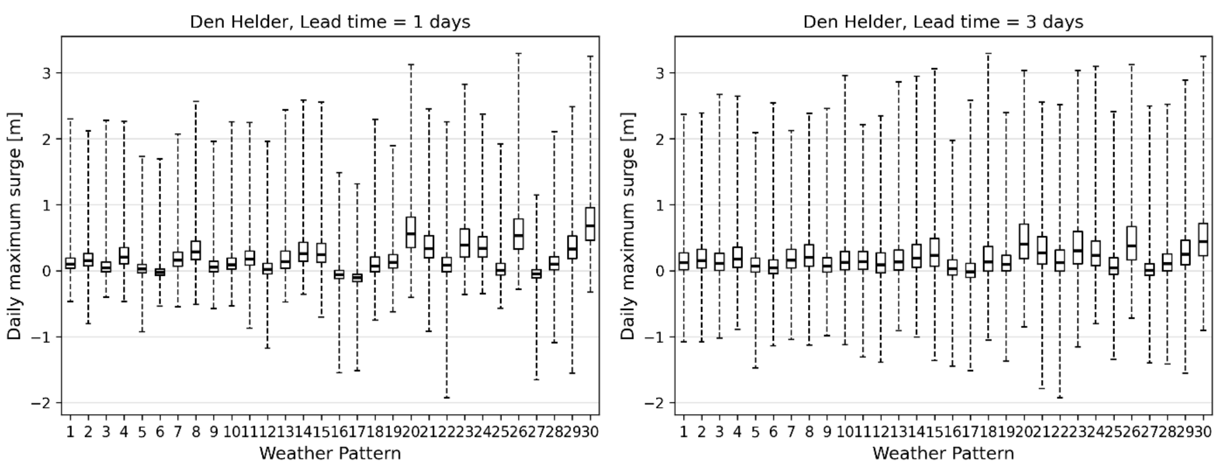
Vlissingen

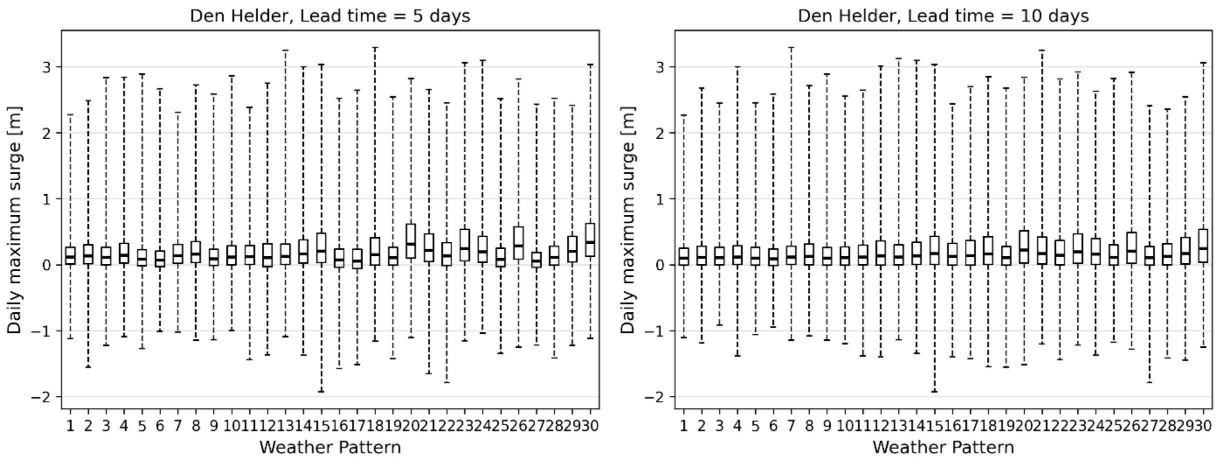


Hoek van Holland

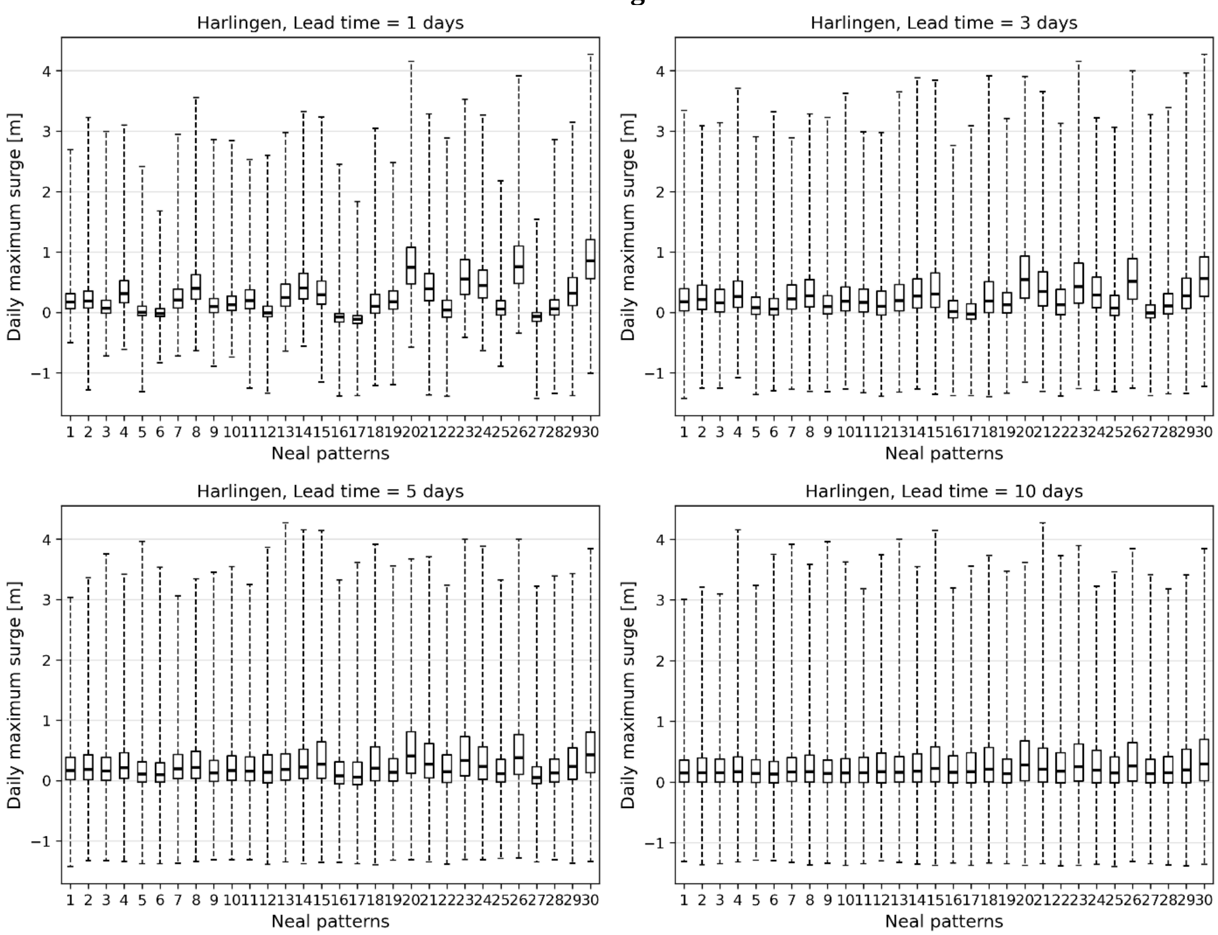


Den Helder

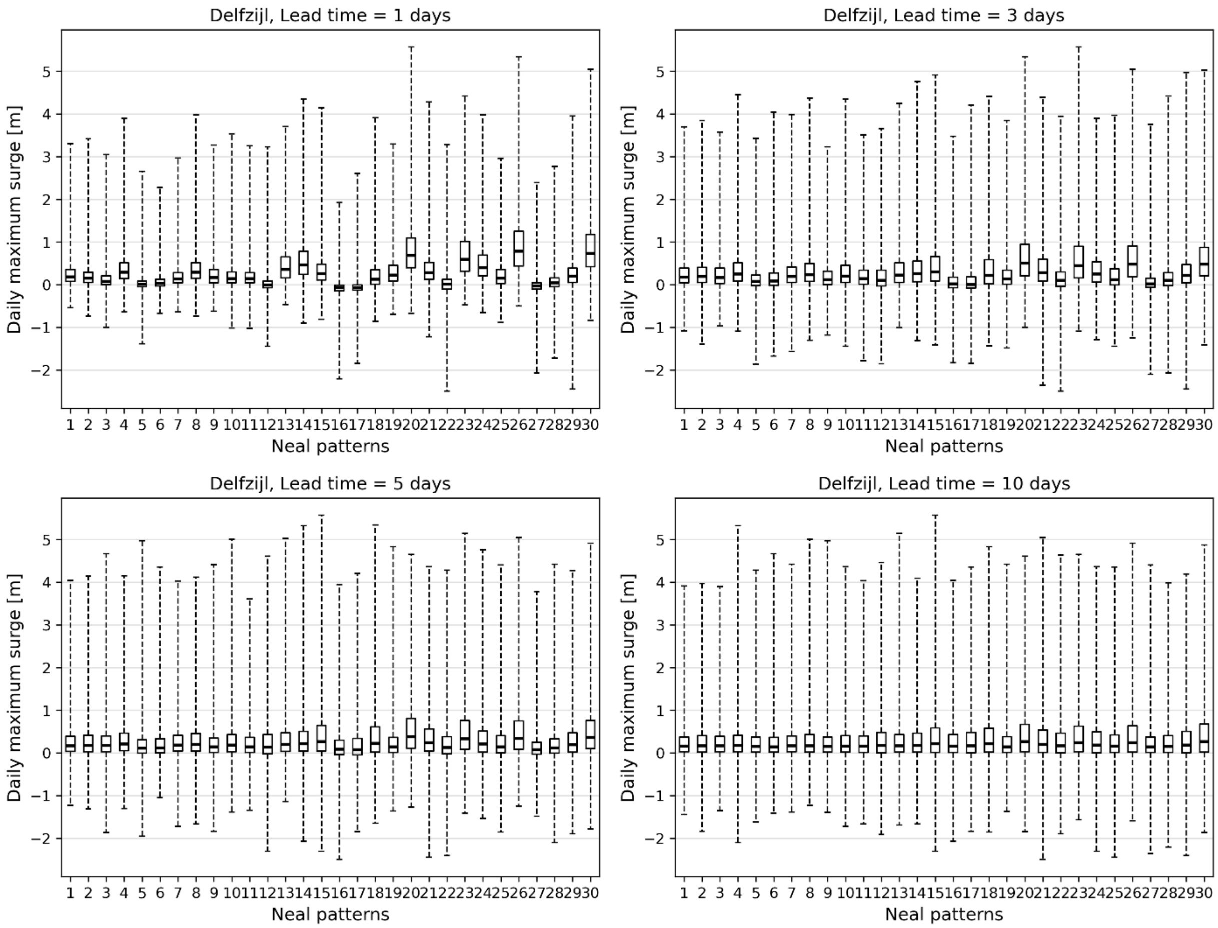



Den Helder, Lead time = 5 days
Den Helder, Lead time = 10 days
Daily maximum surge [m]
Weather Pattern

Harlingen

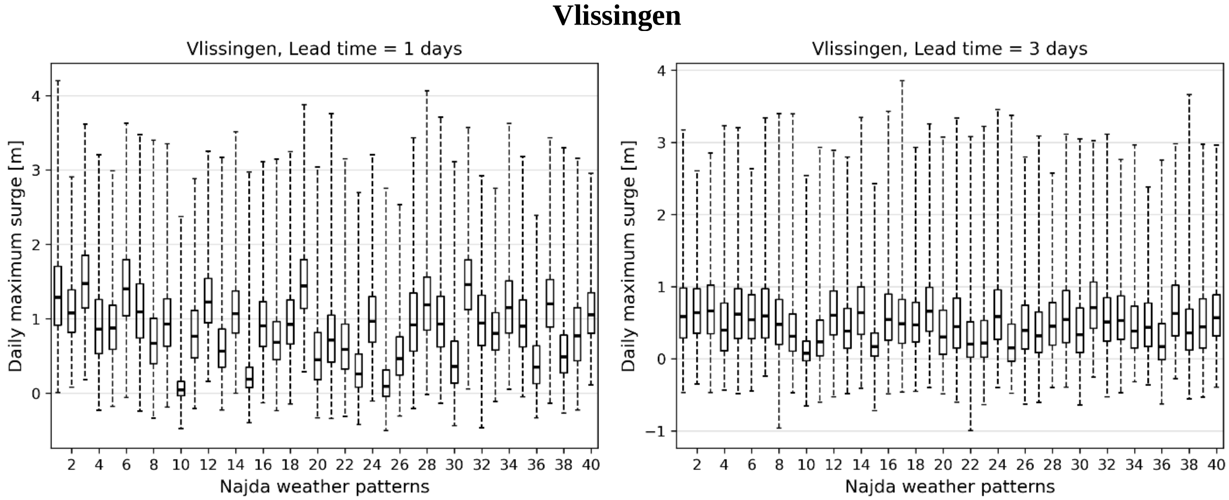

Harlingen, Lead time = 1 days
Harlingen, Lead time = 3 days
Harlingen, Lead time = 5 days
Harlingen, Lead time = 10 days
Daily maximum surge [m]
Neal patterns

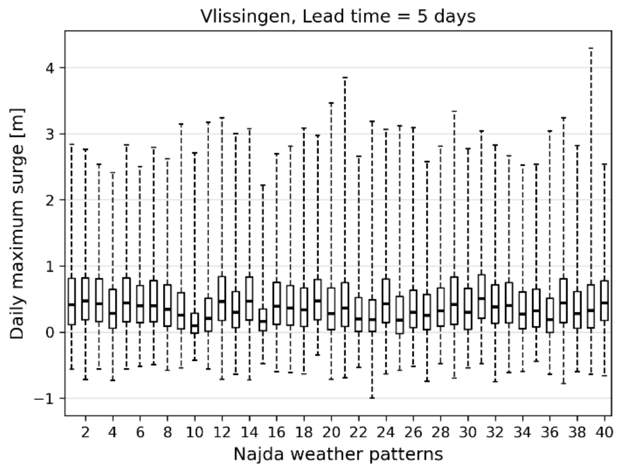
Delfzijl

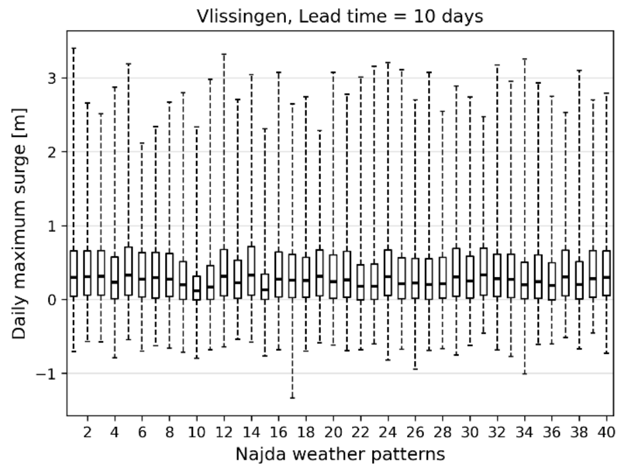


Method A: Surge threshold

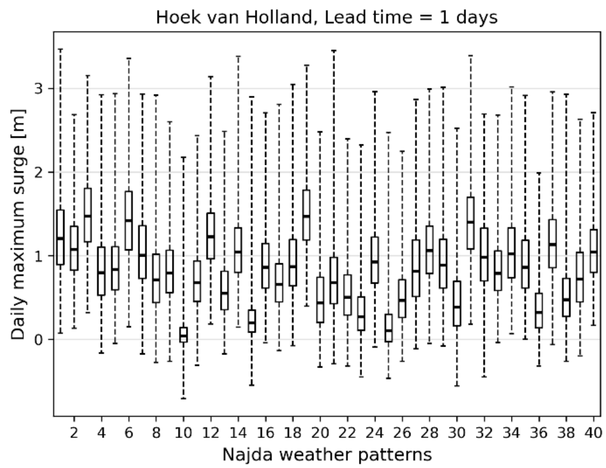
Vlissingen

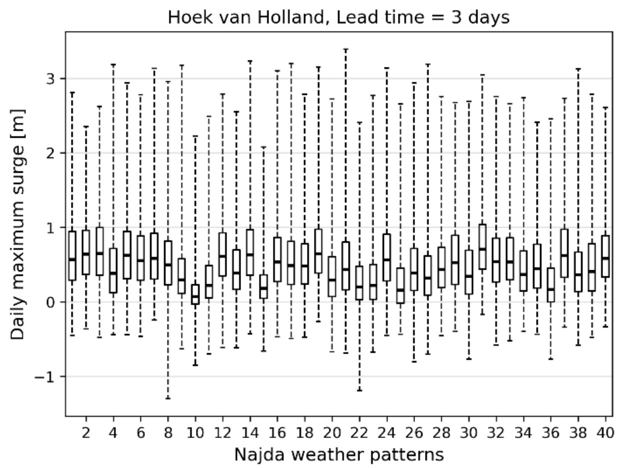


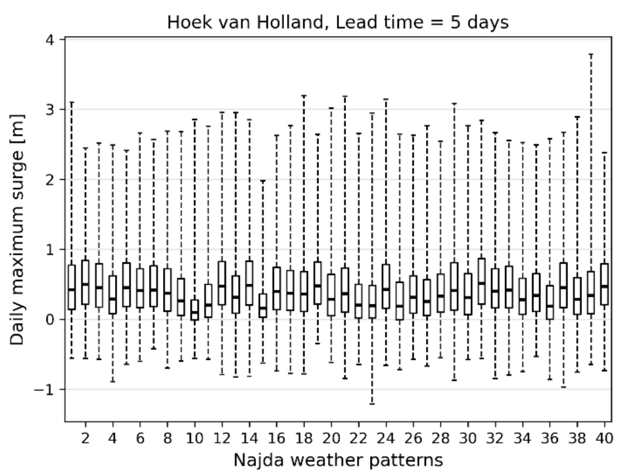


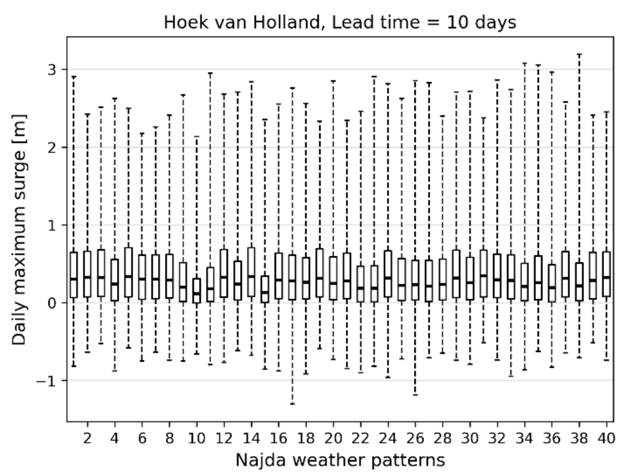


Hoek van Holland

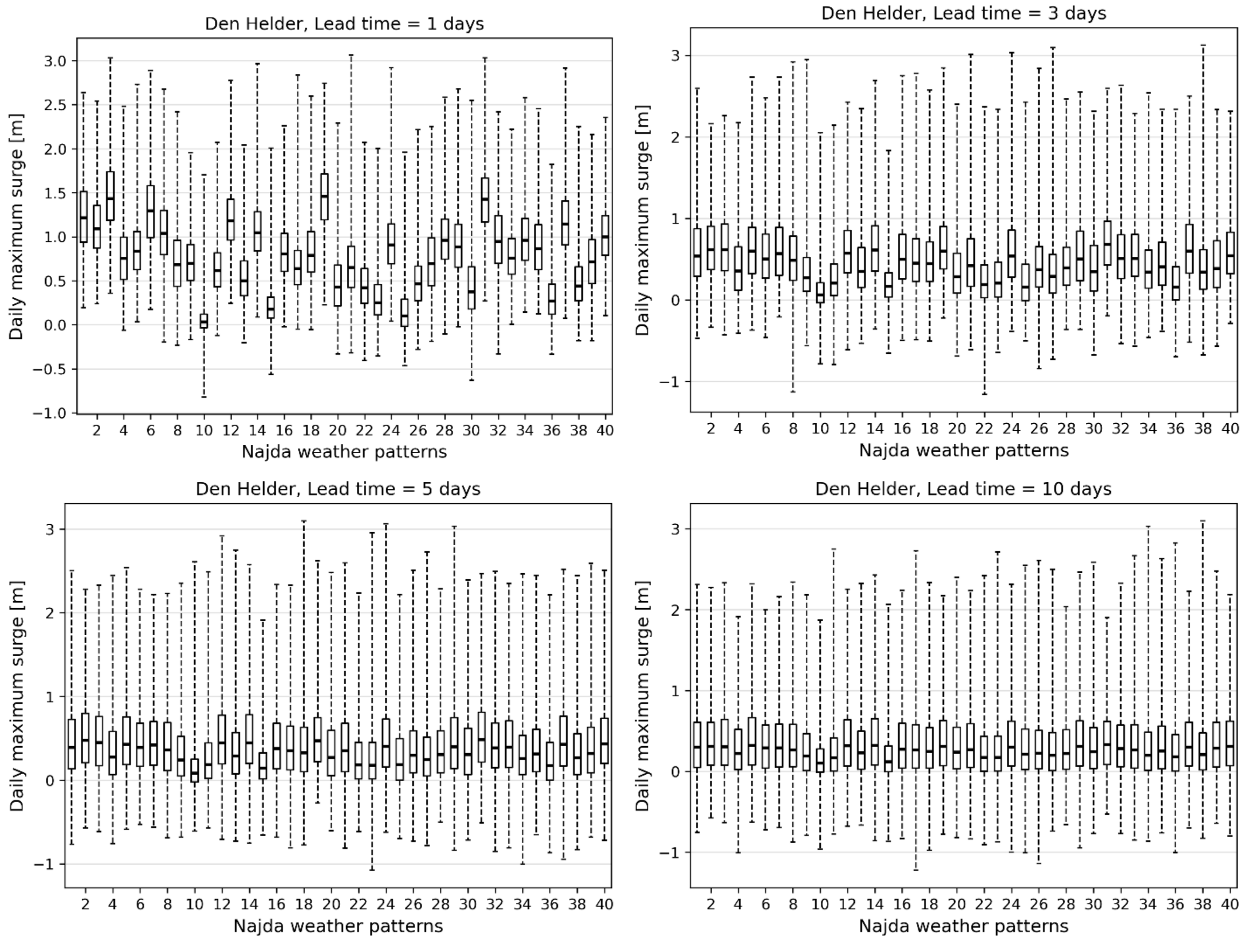




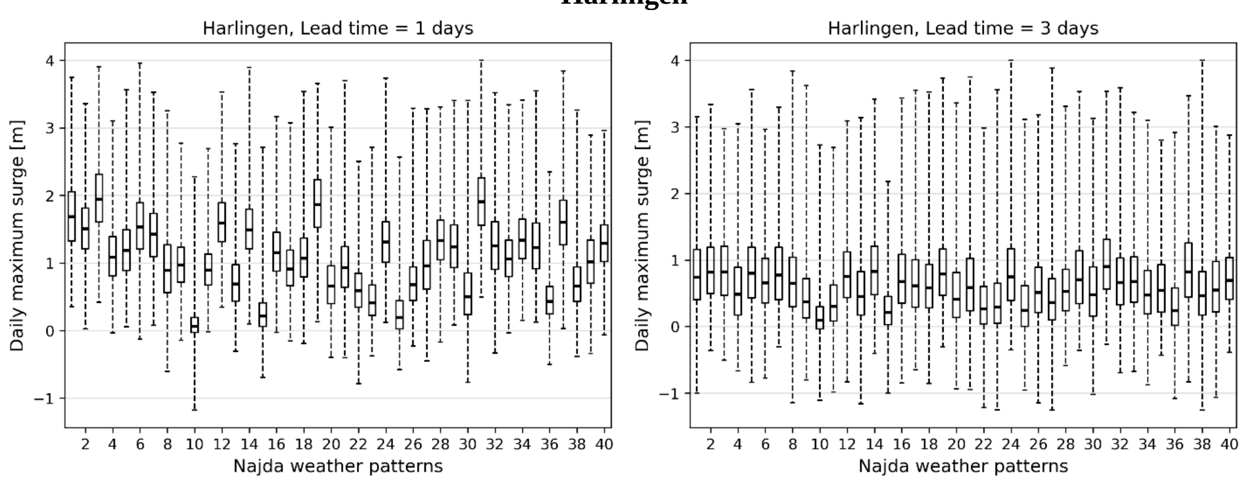


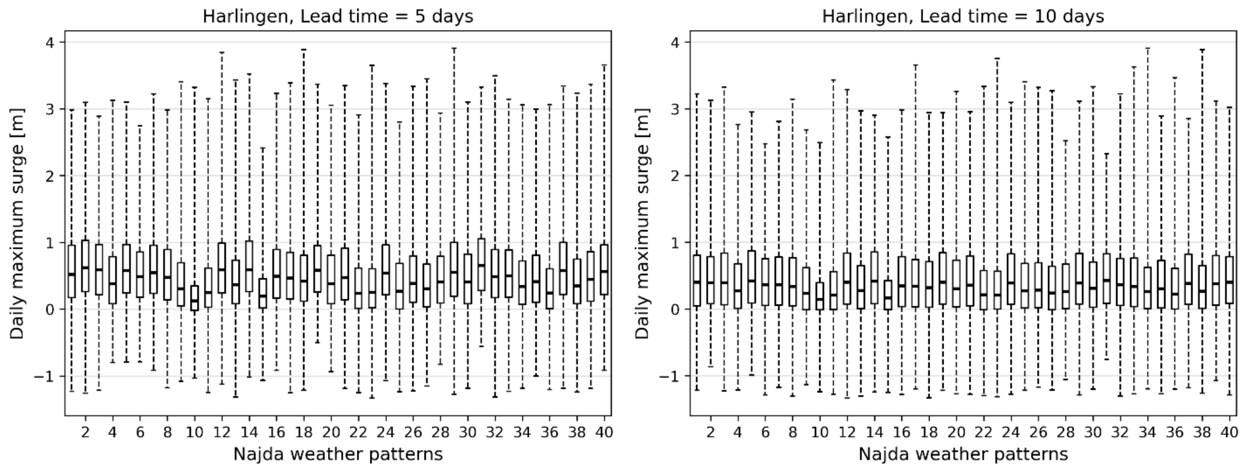


Helder

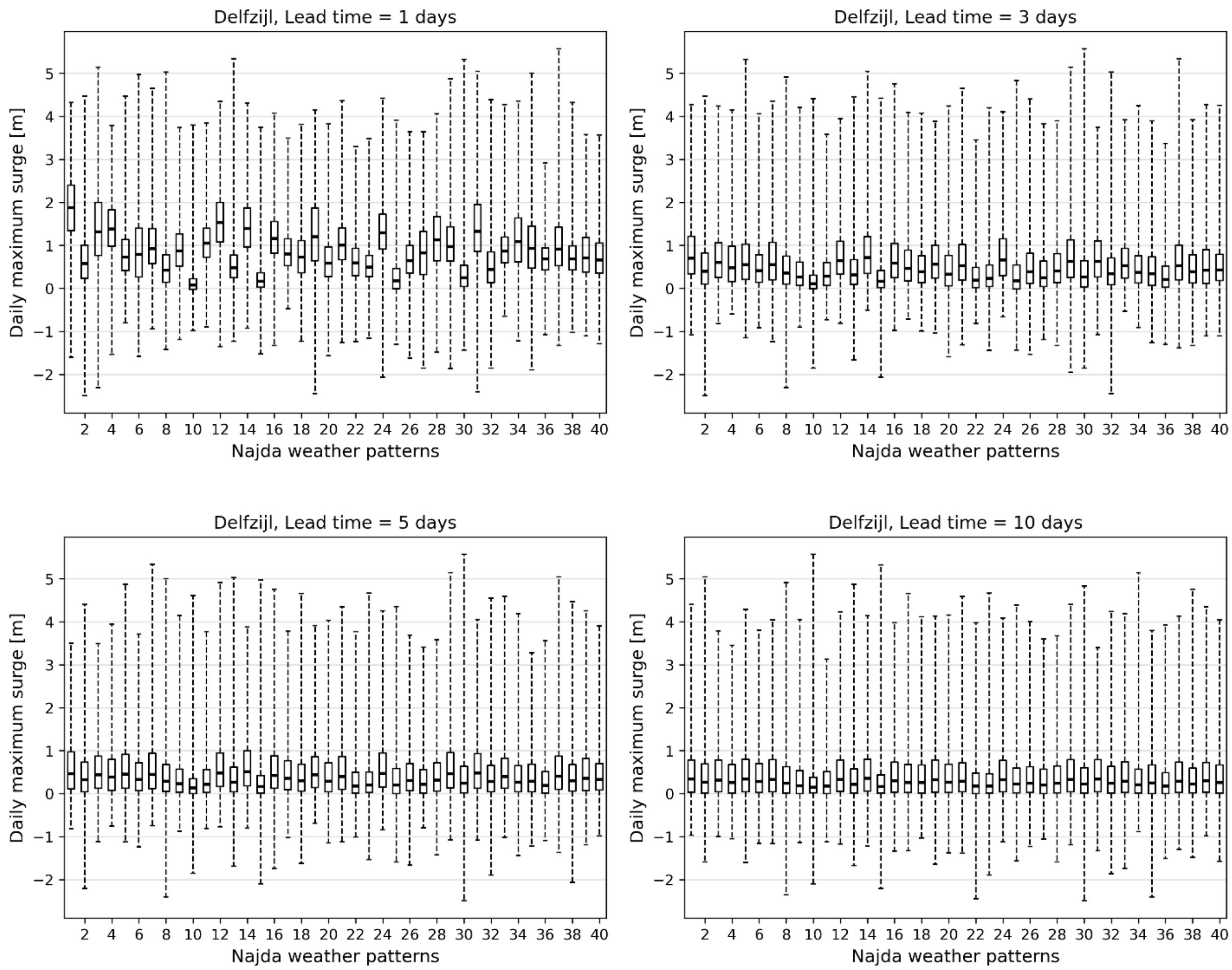


Harlingen



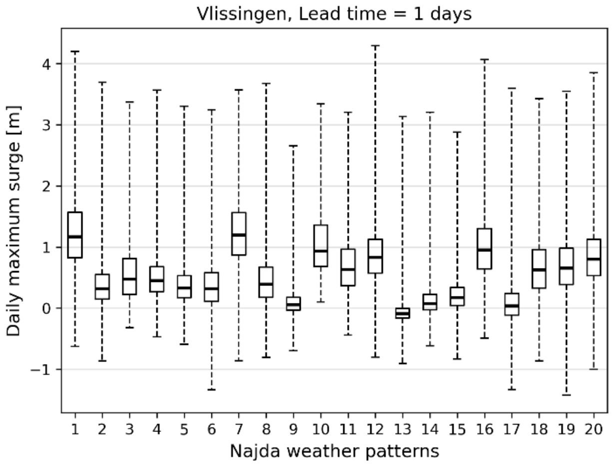

Harlingen, Lead time = 5 days
Harlingen, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

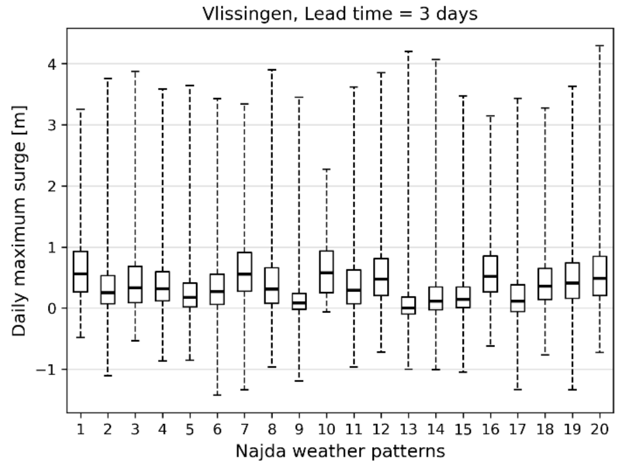
Delfzijl

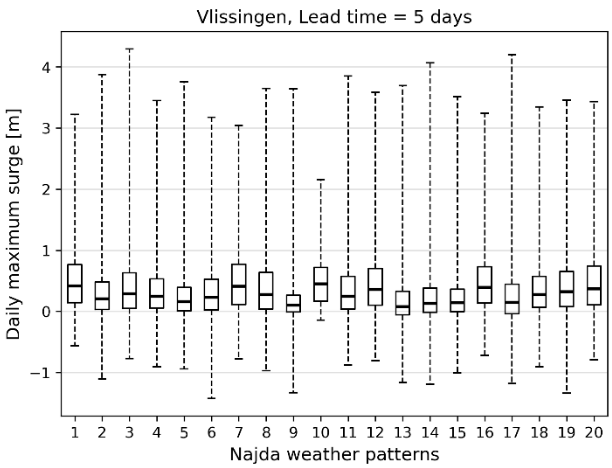

Delfzijl, Lead time = 1 days
Delfzijl, Lead time = 3 days
Delfzijl, Lead time = 5 days
Delfzijl, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

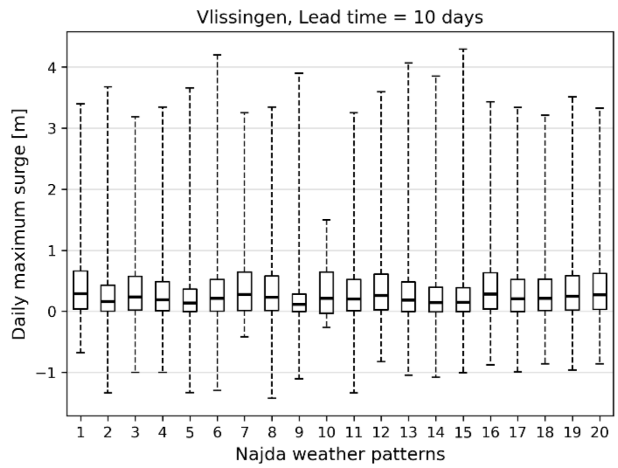
Method B: MDA

Vlissingen

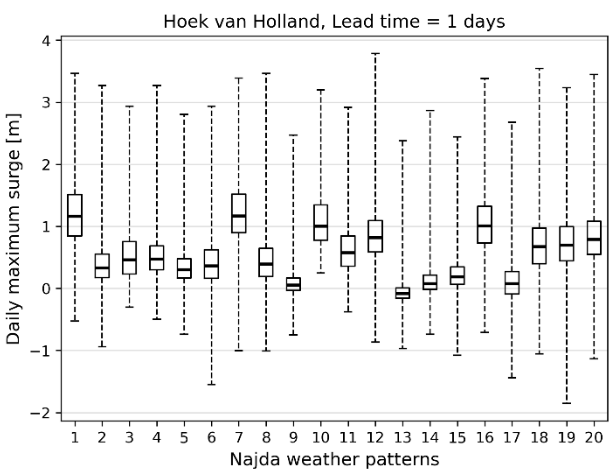


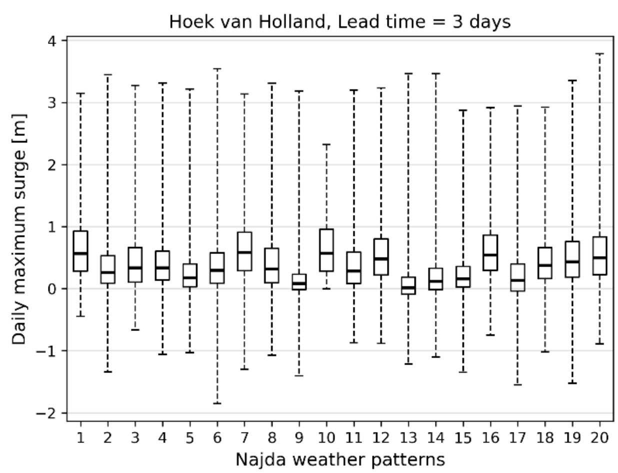


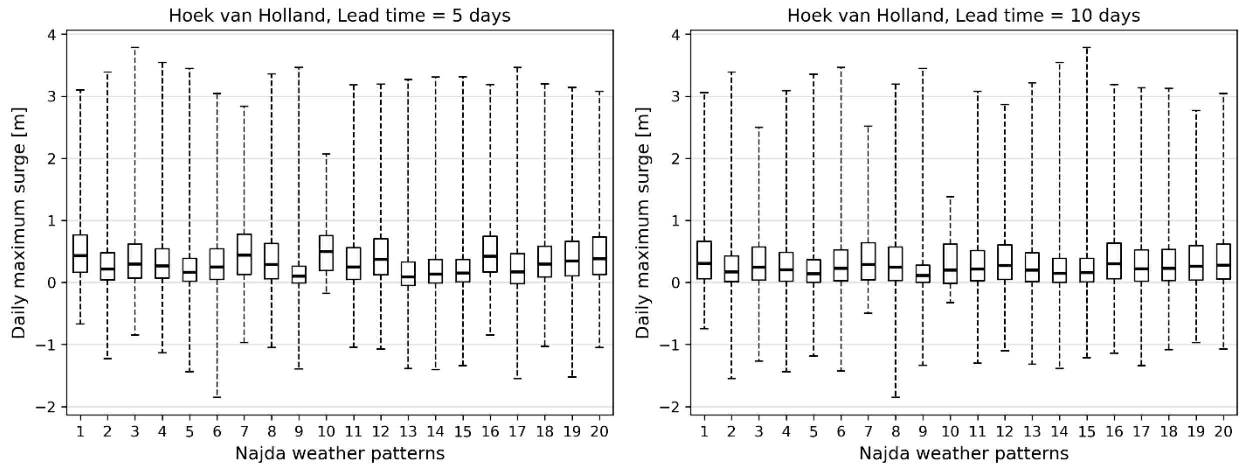




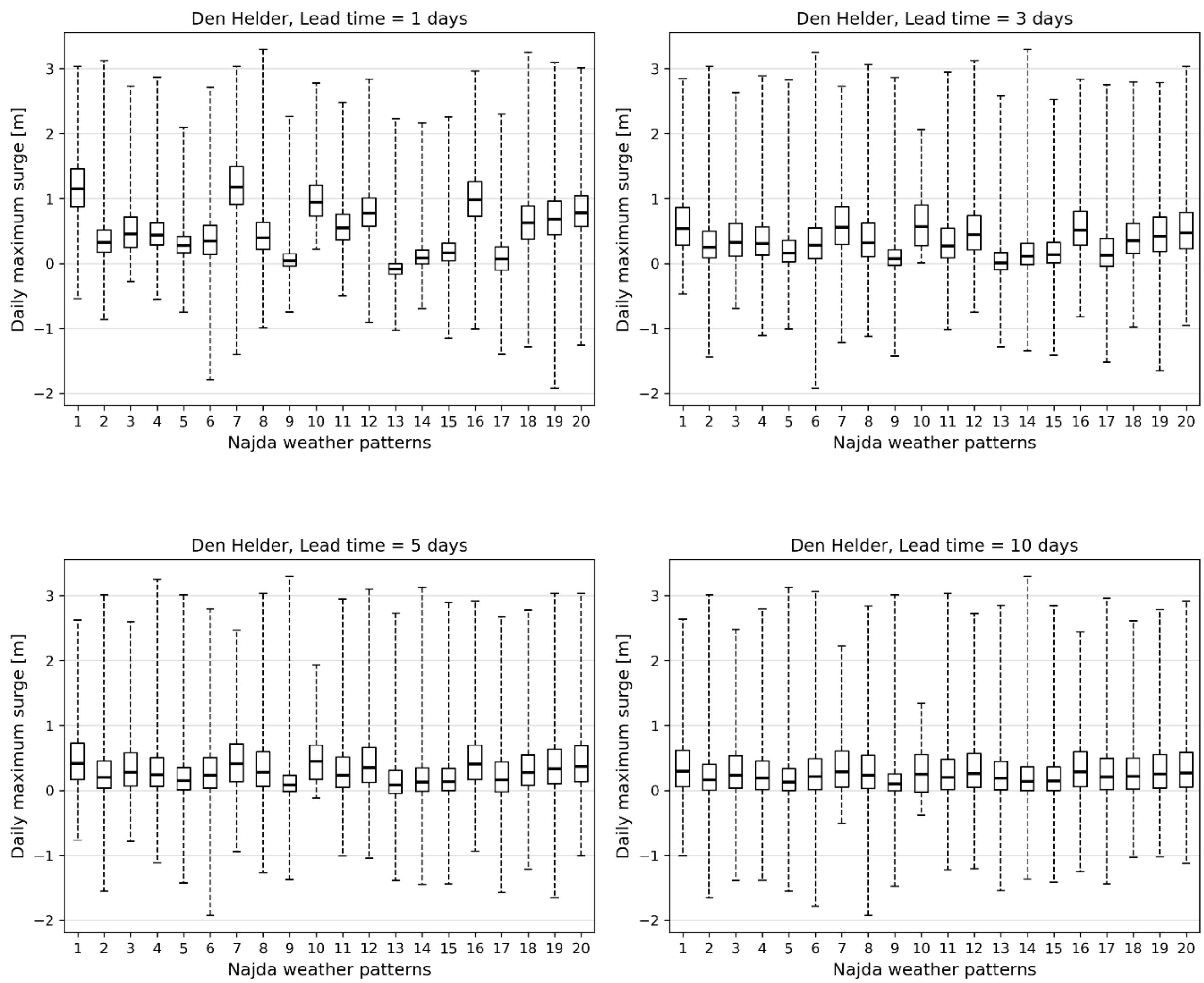
Hoek van Holland



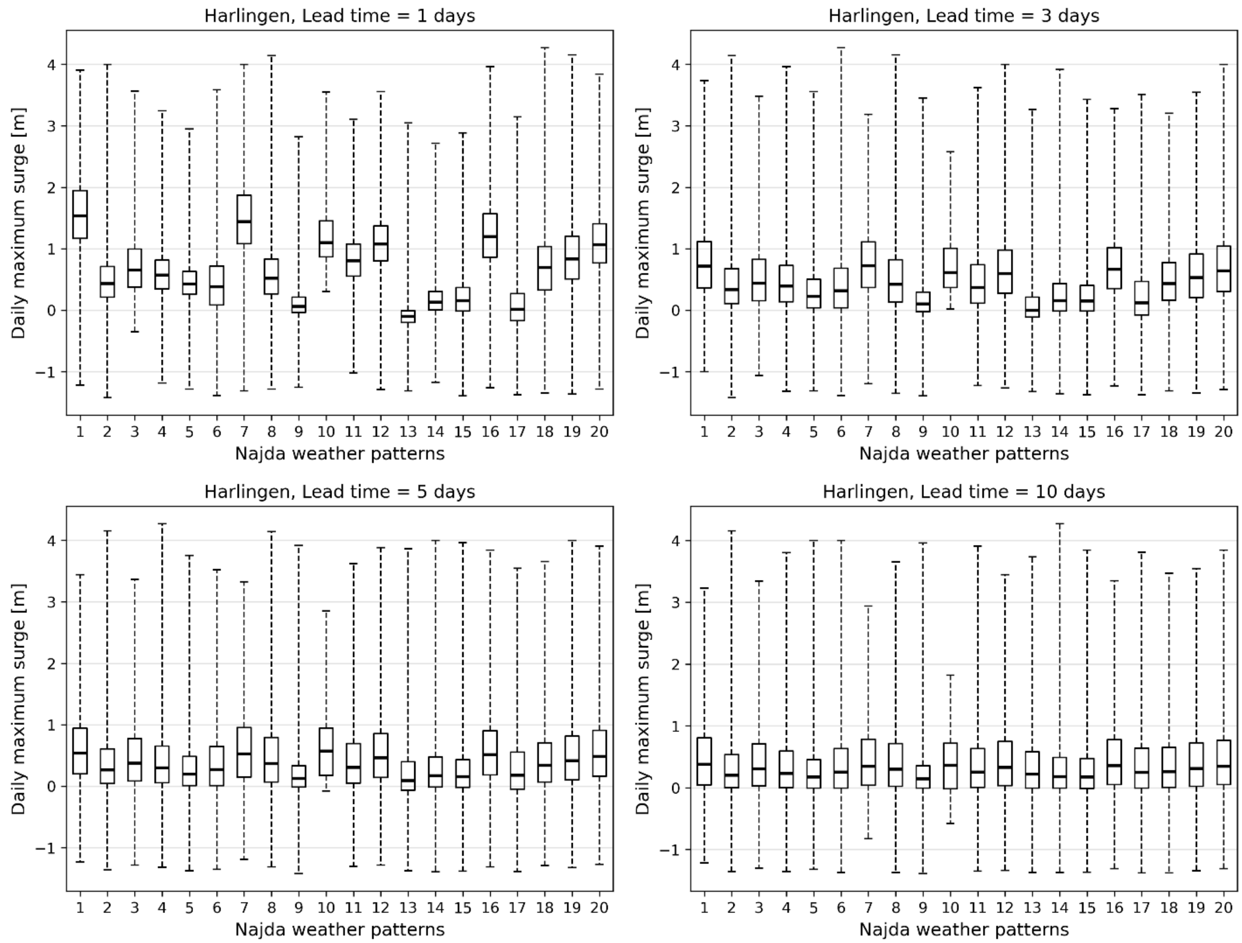



Hoek van Holland, Lead time = 5 days
Hoek van Holland, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

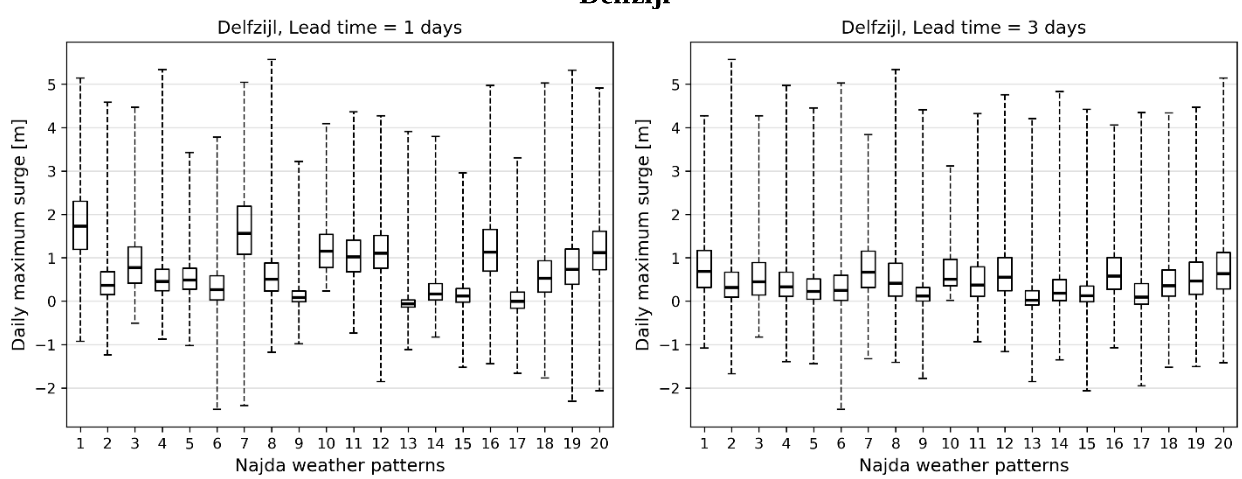
Den Helder

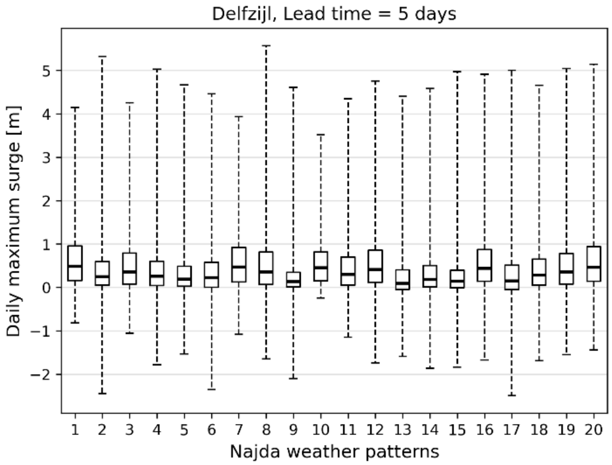

Den Helder, Lead time = 1 days
Den Helder, Lead time = 3 days
Den Helder, Lead time = 5 days
Den Helder, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

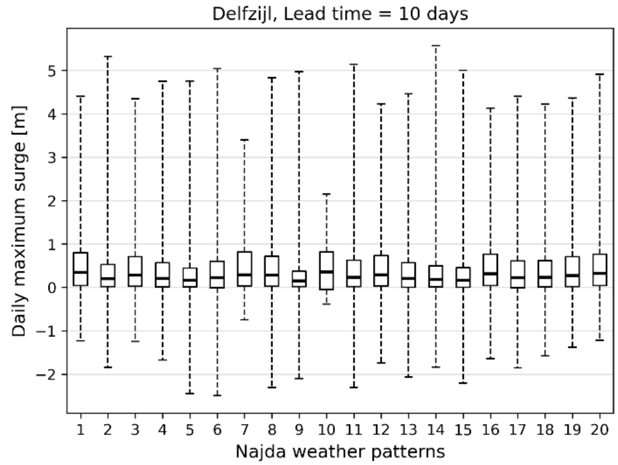
Harlingen



Delfzijl

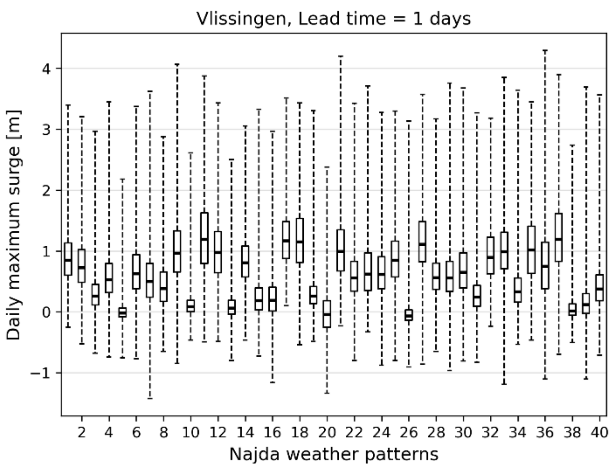


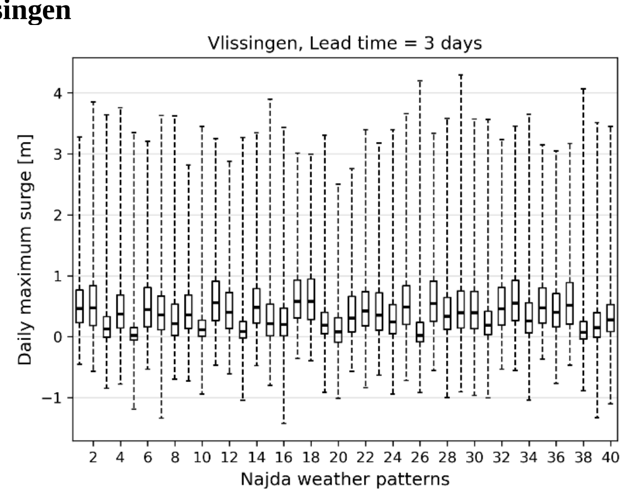


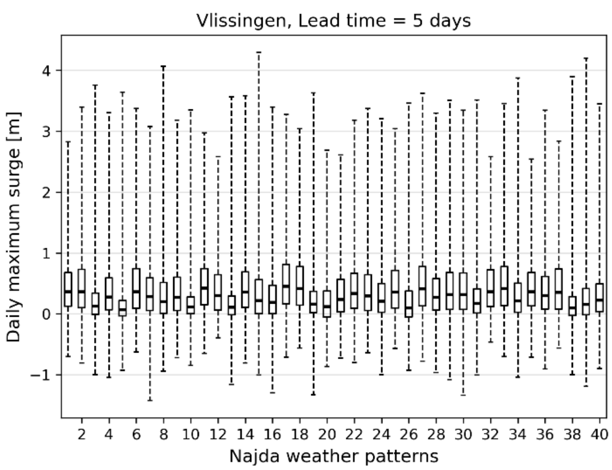


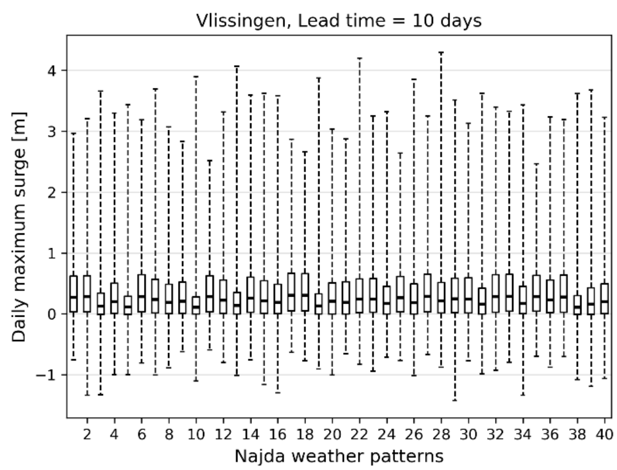
Method C: Stratified sampling

Vlissingen

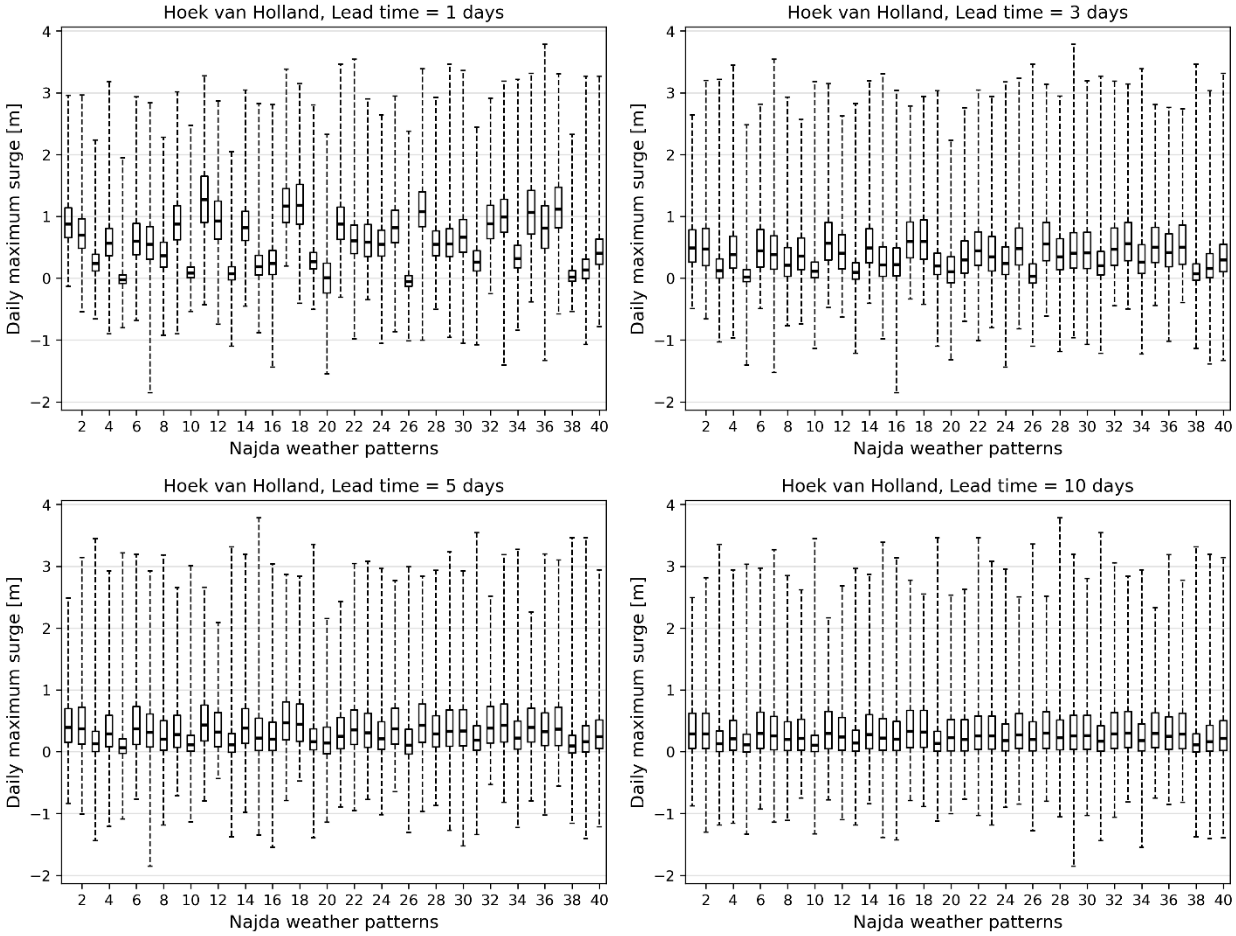




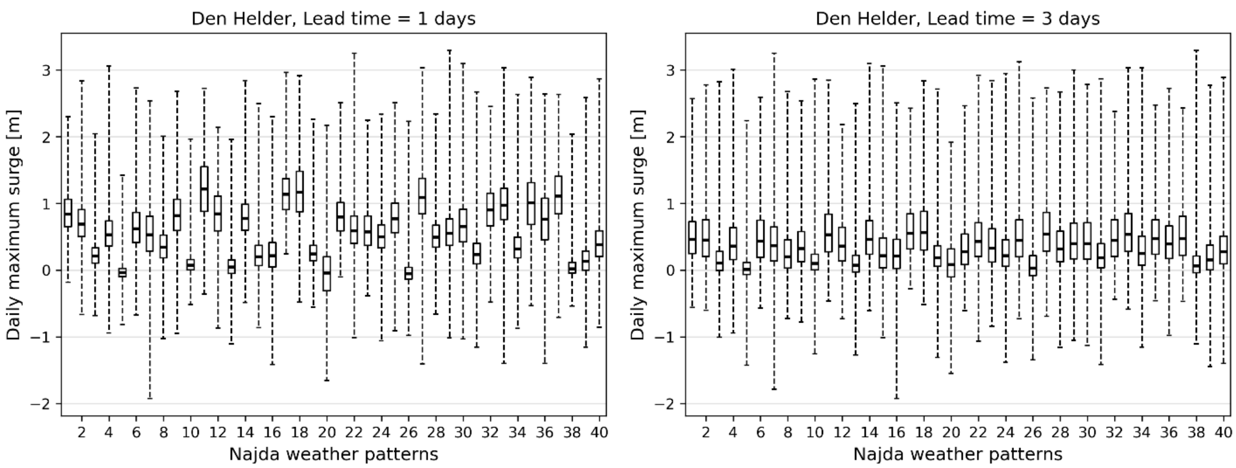


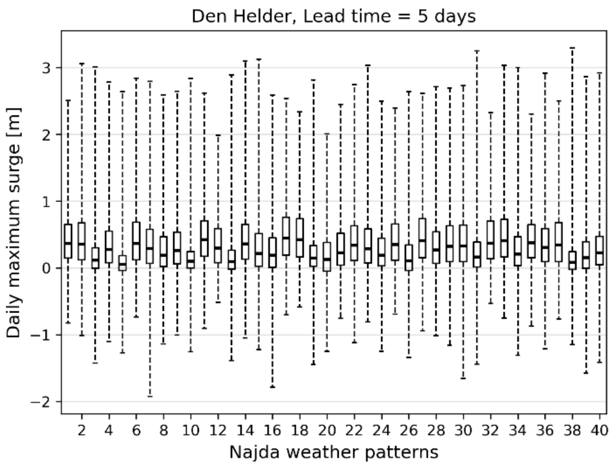


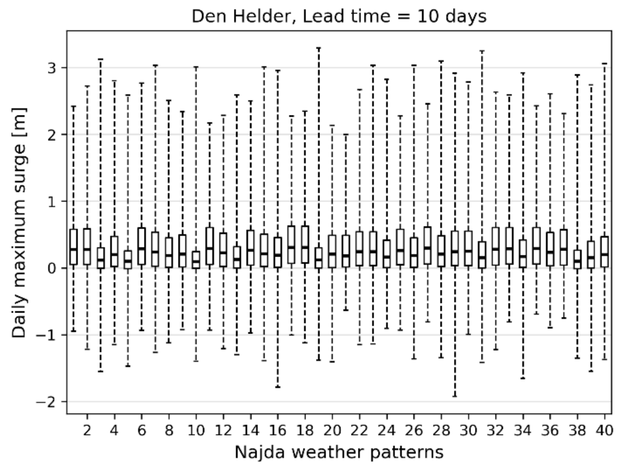
Hoek van Holland



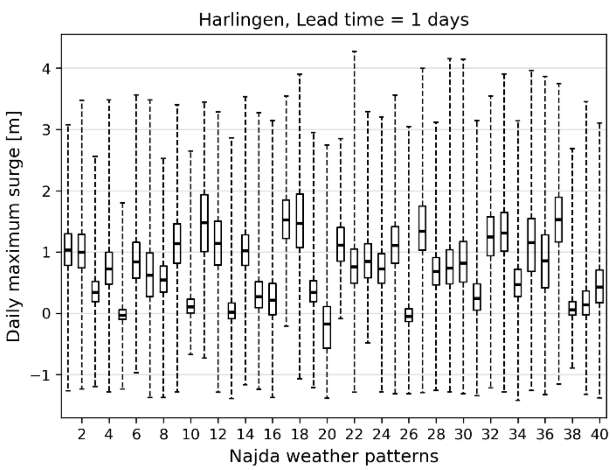
Den Helder

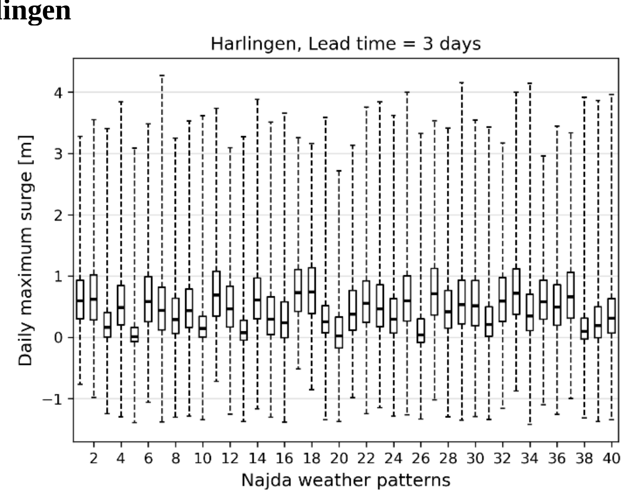


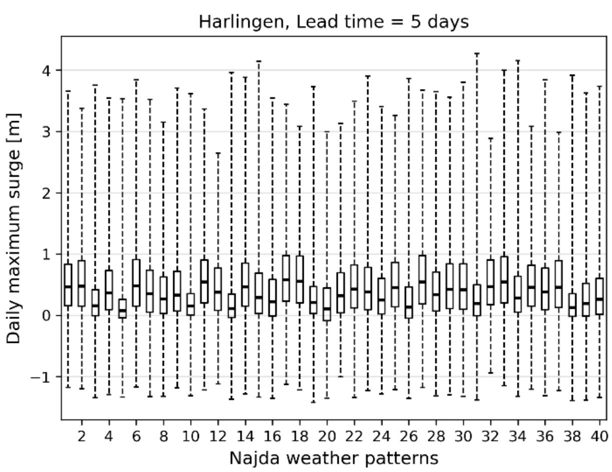

Den Helder, Lead time = 5 days
Daily maximum surge [m]
Najda weather patterns

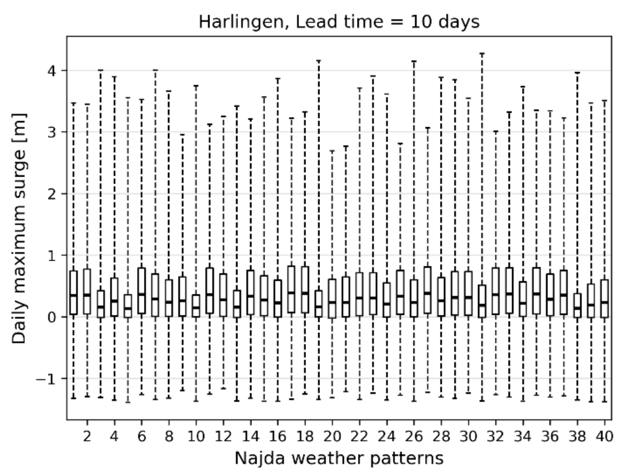

Den Helder, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

Harlingen

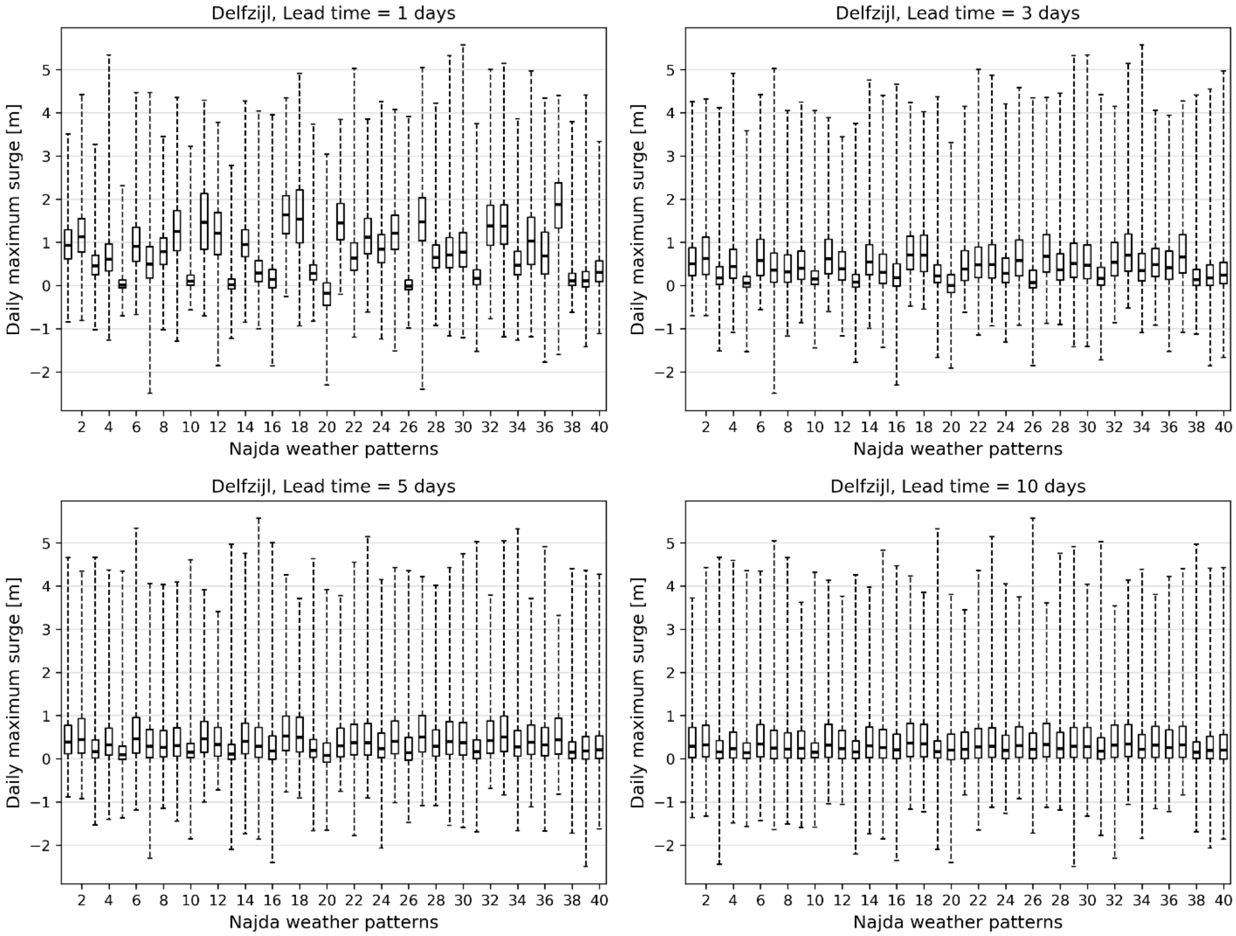

Harlingen, Lead time = 1 days
Daily maximum surge [m]
Najda weather patterns


Harlingen, Lead time = 3 days
Daily maximum surge [m]
Najda weather patterns


Harlingen, Lead time = 5 days
Daily maximum surge [m]
Najda weather patterns


Harlingen, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

Delfzijl



Delfzijl, Lead time = 1 days
Delfzijl, Lead time = 3 days
Delfzijl, Lead time = 5 days
Delfzijl, Lead time = 10 days
Daily maximum surge [m]
Najda weather patterns

Appendix H: Entropy analysis

This appendix presents the results of the entropy analysis conducted for the different clustering methods used in this study. Shannon entropy quantifies the uncertainty within each weather pattern's surge distribution. KL divergence measures the dissimilarity between weather pattern distributions. Figure H-1 presents the Shannon entropy of the surge distributions for Neal's weather patterns (top row), as well as for Methods A, B, and C (subsequent rows). Figure H-2 presents the KL divergence (relative entropy) matrices, showing how different weather patterns compare in terms of surge distribution dissimilarity.

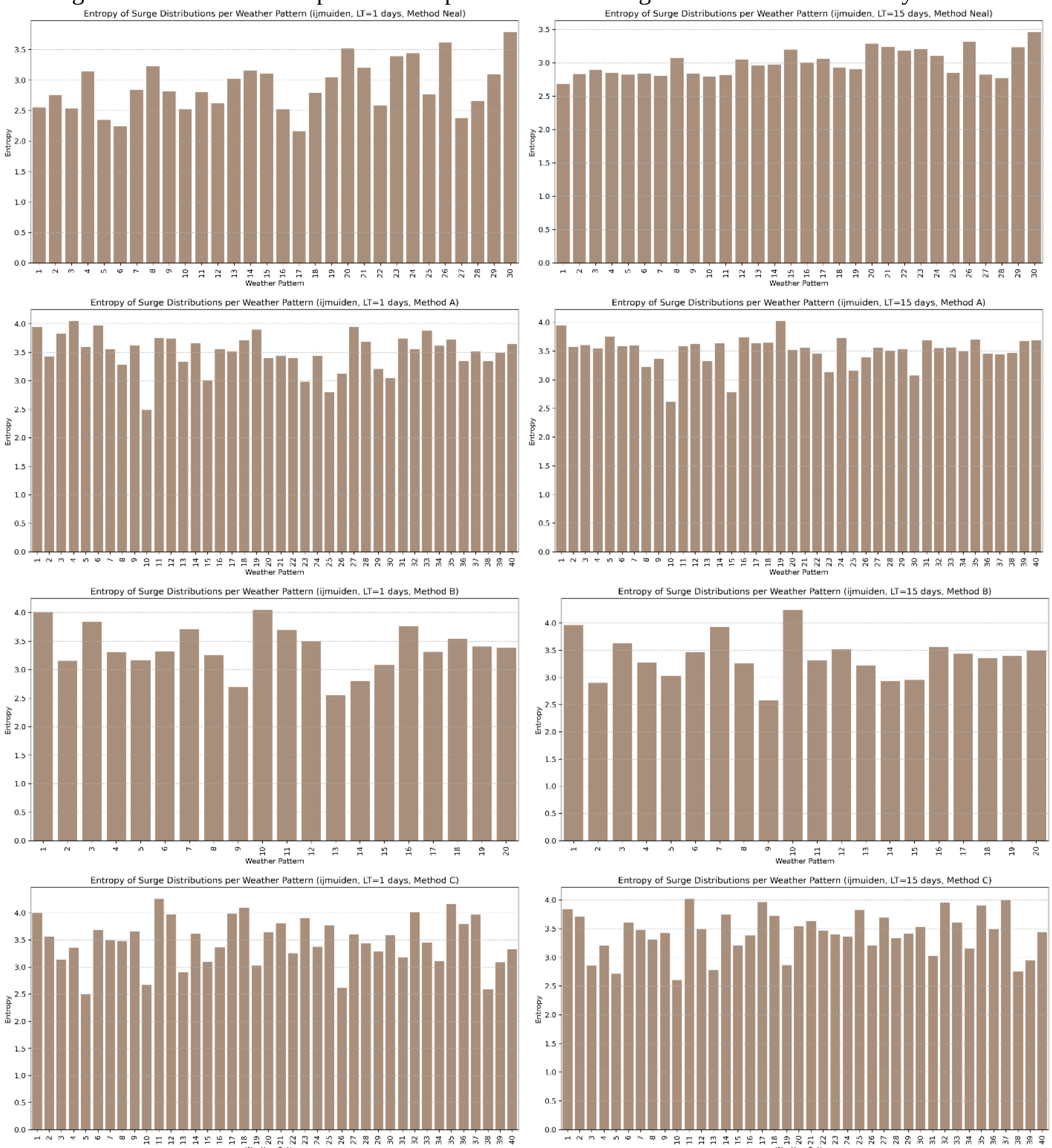
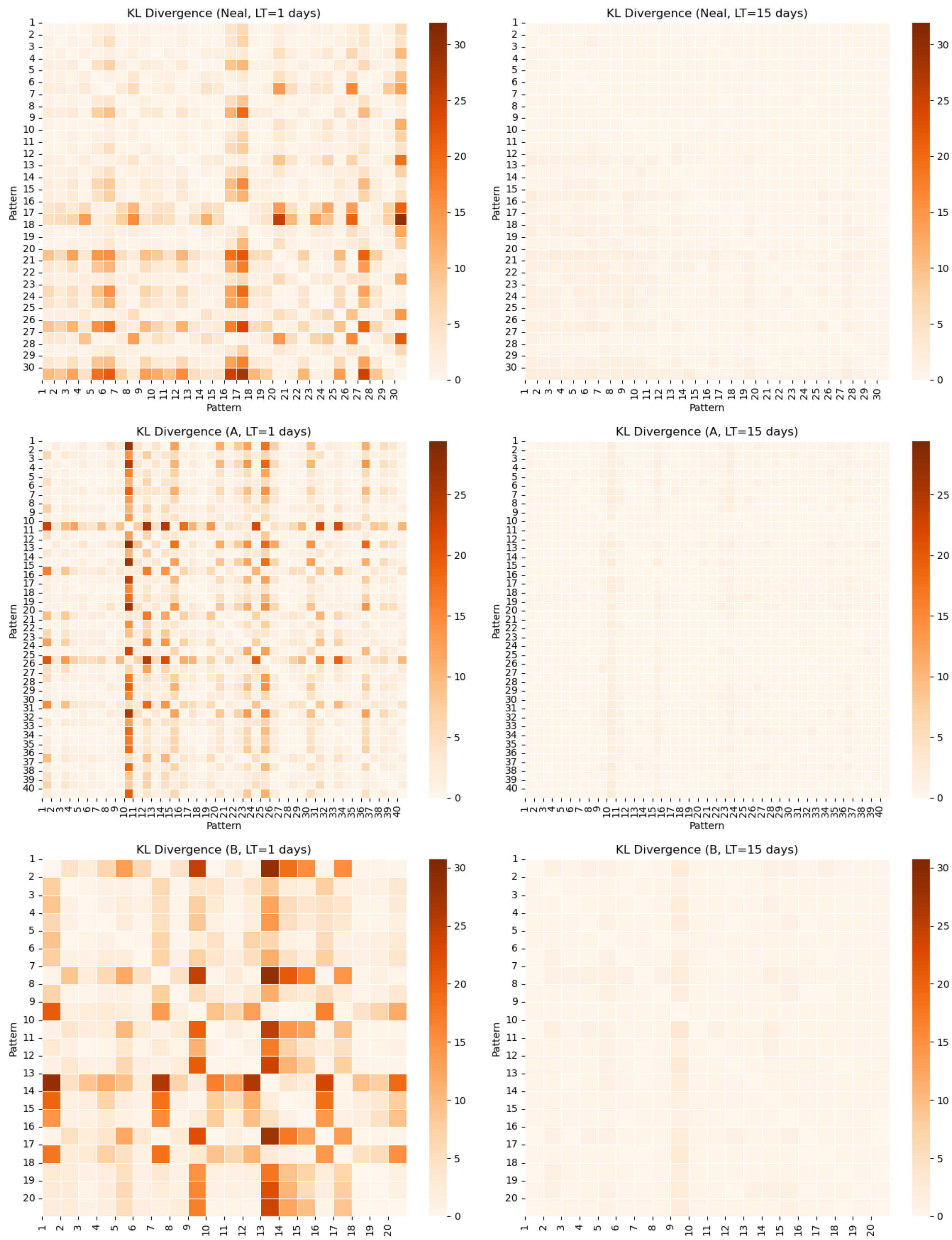


Figure H-1: Shannon entropy values for the surge distributions of each weather pattern across the four clustering methods (Neal, A, B, and C). Higher entropy values indicate greater uncertainty in surge levels within a given pattern.


KL Divergence (Neal, LT=1 days)
KL Divergence (Neal, LT=15 days)
KL Divergence (A, LT=1 days)
KL Divergence (A, LT=15 days)
KL Divergence (B, LT=1 days)
KL Divergence (B, LT=15 days)
Pattern
Pattern

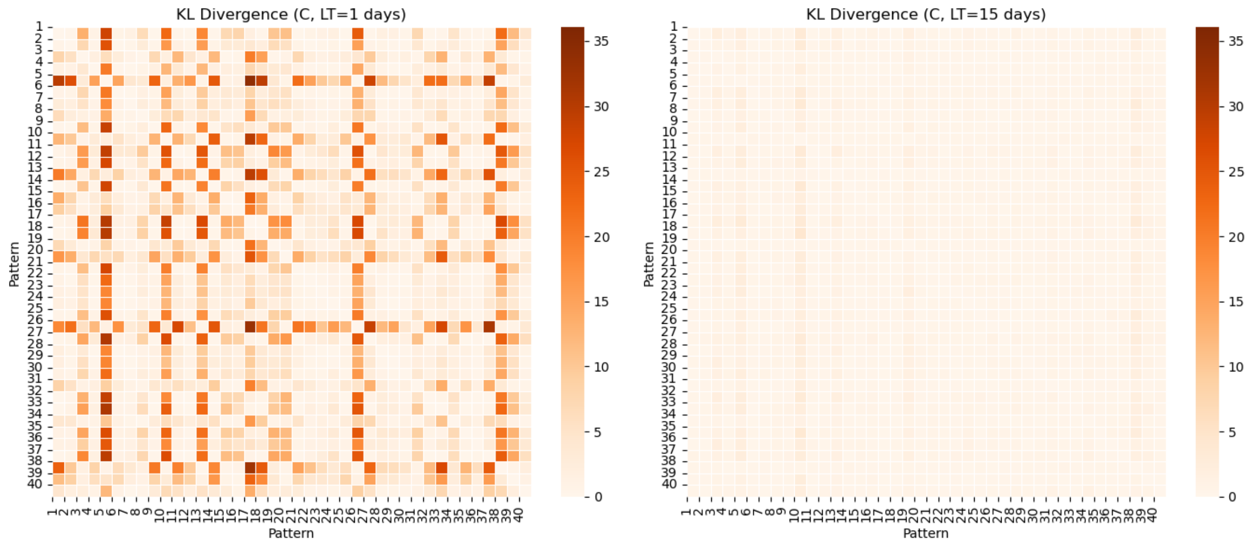


Figure H-2: KL divergence matrices for the surge distributions of each weather pattern across the four clustering methods (Neal, A, B, and C). Higher KL divergence values indicate greater dissimilarity between weather pattern surge distributions.

Appendix I: Trajectory comparison of storm Pia and the representative SEAS5 storm

This appendix illustrates the trajectories of storm Pia and the selected representative SEAS5 storm. The comparison highlights how closely the path of the SEAS5 storm mirrors Pia's trajectory, particularly with regard to its rapid movement across the North Sea.

Figure I-1 illustrates the trajectory of storm Pia, which developed in mid-December 2023. The storm originated near Iceland and moved quickly south-eastward across Northern Europe, impacting the Netherlands.

Figure I-2 shows the trajectory of the selected SEAS5 storm from the ensemble forecast of October 1994, which was used as a proxy for storm Pia. As seen in the figure, the SEAS5 storm follows a similar west-to-east path across the North Sea, with rapid intensification similar to Pia. Both storms exhibit fast progression across the region, reinforcing the SEAS5 storm's suitability as a proxy for this analysis.

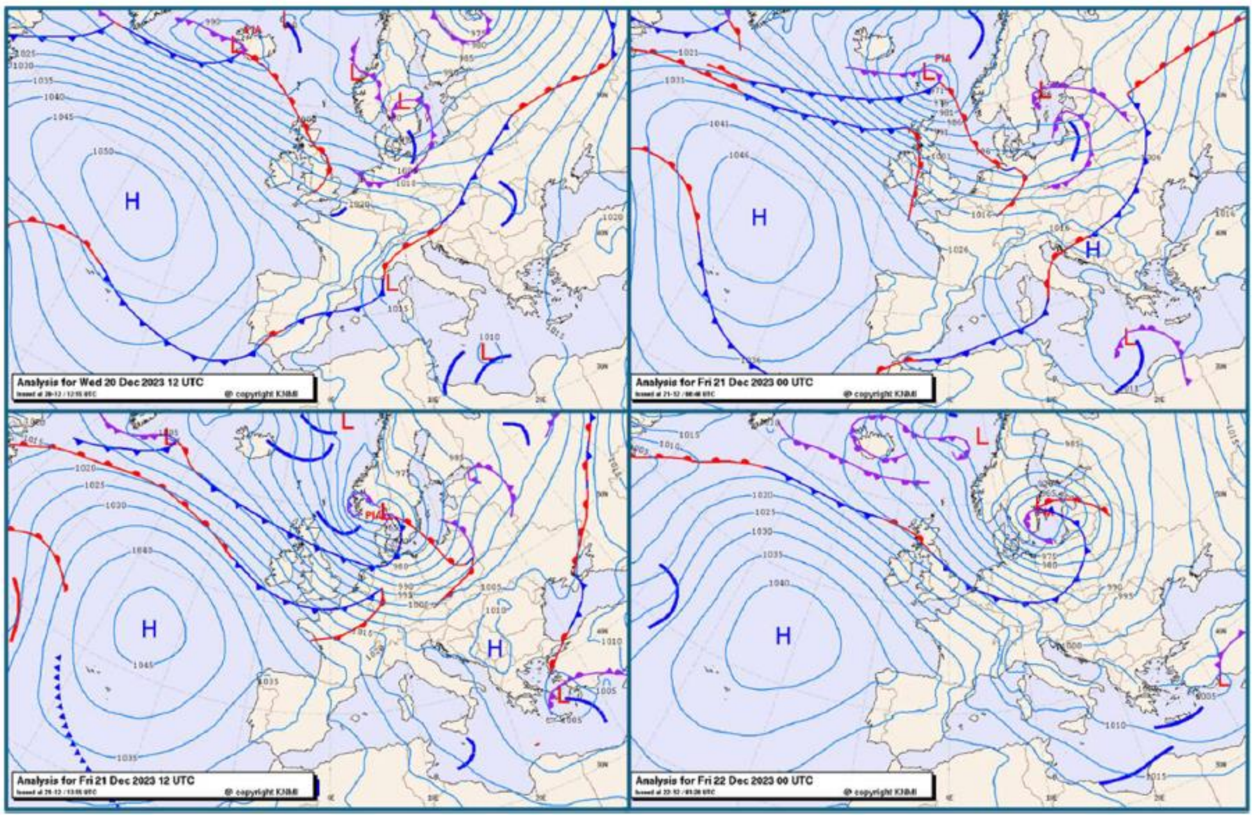


Figure I-1: Trajectory of storm Pia (December 2023), showing its rapid movement from Iceland towards the Baltic Sea. Taken from (Zijderveld, et al., 2024).

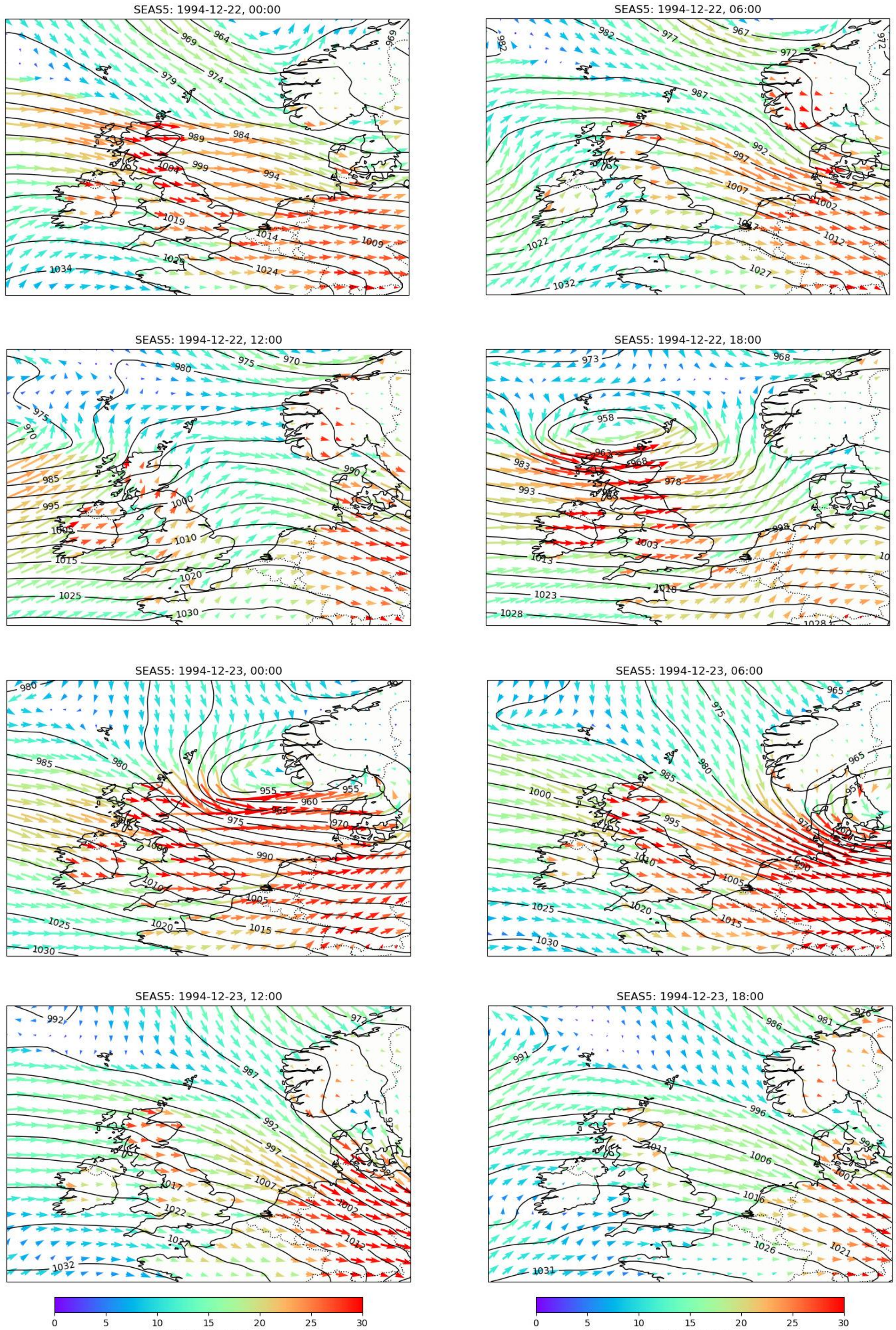
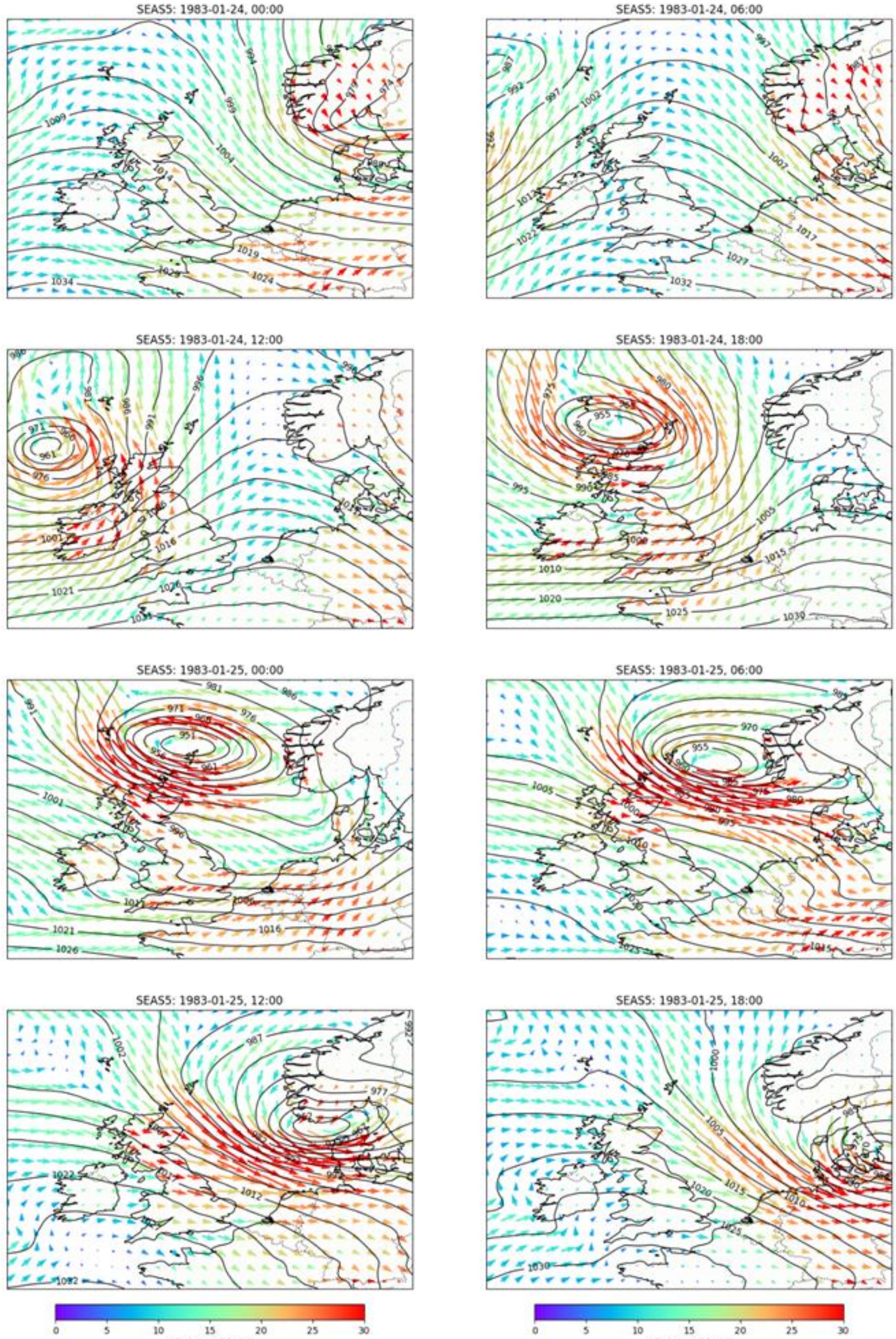


Figure I-2: Trajectory of the representative SEAS5 storm (22-23 December 1994), showing its similar west-to-east path and rapid movement across the North Sea.

Appendix J: Storm evolution in the considered domain

This appendix provides a detailed visualisation of how storms develop and enter the SEAS5 model domain before leading to extreme surge events. As discussed in Section 4.4, an analysis was conducted on 22 days with maximum daily surges exceeding 4.5 m, examining the mean sea level pressure (MSLP) and wind intensity fields 15 days prior to each surge event. The analysis revealed that for all cases, the responsible storm system only entered the domain approximately 1–2 days before the high-surge event occurred.

To illustrate this pattern, this appendix presents an example of one such storm. The contour lines in the figures represent the MSLP fields, while the arrows indicate wind speed and direction. For the high-surge event on January 25, 1983, at 18:00, the storm system entered the SEAS5 domain on January 24, 1983, at 06:00, meaning only 36 hours elapsed between the storm's first appearance in the domain and the peak surge. These visuals support the conclusion that the limited spatial domain may restrict the ability to predict extreme surge events well in advance.


SEAS5: 1983-01-24, 00:00
SEAS5: 1983-01-24, 06:00
SEAS5: 1983-01-24, 12:00
SEAS5: 1983-01-24, 18:00
SEAS5: 1983-01-25, 00:00
SEAS5: 1983-01-25, 06:00
SEAS5: 1983-01-25, 12:00
SEAS5: 1983-01-25, 18:00
0 5 10 15 20 25 30
Wind Speed (m/s)
0 5 10 15 20 25 30
Wind Speed (m/s)