



Source: <https://studyinginswitzerland.com/swiss-movies/>

# Predicting next month's average temperature for energy trading

Bachelor thesis

Degree programme:

Industrial Engineering and Management Science BSc

Author:

Joël Tauss

1<sup>st</sup> Supervisor:

Prof. Dr. Angela Meyer

2<sup>nd</sup> Supervisor:

Prof. Dr. Stefan Grösser

Project partner:

Bern University of Applied Sciences

Date:

2023-06-28

## Abstract

This thesis delves into the topic of temperature prediction, analysis of the polar vortex weather phenomenon, and the correlations between the outdoor temperature and the gas prices. The primary objective was to create a temperature prediction model by finding ways to utilize custom polar vortex indexes and data in combination with other metrics, such as weather indexes, like the Southern Oscillation and Antarctic Oscillation index. Because this model specifically caters to energy trading (gas) it was created in the form of classifying two lead times, into a below (0) and above (1) average temperature category for the given days: t1 - 0 to 14 days ahead; t2 - 15 to 28 days ahead. Therefore, the model predicts if the heating energy which will be needed is more or less than the mean historical amount.

The polar vortex is a circular wind over the North Pole region, occurring at altitudes of around 15 to 50 km. It is stronger during the winter season and weaker during summertime. These changes can influence Europe's and North America's outdoor temperatures. In order to create two custom weather indexes out of the polar vortex, one for temperature and one for wind speeds, multiple approaches were tested, resulting in two different algorithms for each index. Additionally, image clustering was applied to propose an alternative to the index-based approach.

Three types of models were tested for the classification: a random forest, a multi-layer perceptron, and a recurrent neural network. After thorough optimization, the random forest performed the best with an accuracy of 0.665 for t1 and 0.660 for t2. Hence proofing, that it possesses some predictive skill.

Although the researched literature stated that a connection between the outdoor temperature and the gas spot price during the heating months (October to April) exists, this correlation could not be found within the used data.

# Contents

1	Introduction	5
1.1	Context and Relevance	5
1.2	Goals and Research Question	6
2	Literature	8
2.1	Influences on Gas Prices	8
2.2	Weather Forecasting Models	8
2.3	Temperature Forecasting	9
2.4	Temperature Anomaly Detection	10
2.5	Weather Phenomena and Effects	11
2.6	Literature Summary	13
3	Methods	15
3.1	Tools	15
3.2	Research Design Overview (CRISP-DM)	15
3.3	Business Understanding	16
3.4	Data Understanding	16
3.5	Data Preparation	19
3.6	Modelling - Temperature Forecasting	22
3.7	Modelling – Polar Vortex Clustering	23
3.8	Evaluation	23
3.9	Gas Price and Outdoor Temperature Correlations	24
4	Results	26
4.1	Data Understanding – Data Gathering	26
4.2	Data understanding – Data exploration	29
4.3	Data Preparation	30
4.4	Modelling – Classification Models	39
4.4.1	Data Preparation	39
4.4.2	Random Forest	39
4.4.3	Nu-support Vector Machine	41
4.4.4	Multi-layer Perceptron	41
4.4.5	Recurrent Neural Network	43
4.5	Modelling - Top Model Optimization (RF)	44
4.6	Modelling – Regression Model (MLP)	45
4.7	Modelling – Polar Vortex Clustering	46
4.8	Gas Price and Outdoor Temperature Correlations	47
5	Discussion	50
5.1	Polar Vortex Analysis	50
5.2	Temperature Predictions and Machine Learning	50
5.3	Gas Price and Outdoor Temperature Correlations	51
5.4	Additional Findings	51
5.5	Limitations	51
6	Conclusion	52
6.1	Summary	52
6.2	Further Research	52
7	List of Illustrations	54
8	List of Tables	55
9	List of Equations	55
10	References	56
11	Appendix	60
11.1	Thesis to jupyter notebook Chapter References	60
11.2	Referenced Plots and Graphs	60
11.3	Clustering Sample	68

11.4 Network Architectures	75
11.5 Robust Peak Detection (z-score)	77
12 Declaration of Authorship	78

# 1 Introduction

## 1.1 Context and Relevance

Weather forecasting is an important topic in many different fields of study, industries, and private lives. The gained information of forecasts is applied in various ways, some of which can have a big impact on financial budgets, world markets and business planning. Industries such as agriculture, infrastructure and building constructions, and the energy sector make heavy use of weather data and forecasting, to name a few. [1]

A wide range of forecasting data is already available to the public, may this be various online resources, official weather prognoses from meteorology offices all around the world or independent research papers. But most of the lead times and data are only accurate for up to 10 days. After that time mark, the accuracy drops to about 50%. [2]

Because of the difficulty for long term forecasting weather and especially temperature, this thesis explores this topic as its focal point. The focus of the practical application for the forecasting lies in the energy industry, heating demand forecasting in particular. The proposed type of model is a classification-based approach, which was set because of framework conditions of the project. The classification is applied in such a way, that the temperature datapoints are either labelled as above or below average. The base for this comparison will be a historical value. A regression-based approach is also valid, but not the main goal of this thesis. With a classification approach, the energy traders would be able to assess, if a certain time period will be particularly warm or cold. Such a prediction could then be used in conjunction with past data and consumptions to make an educated guess about the near-future energy demand. Or to put it differently: The model predicts if the heating energy which will be needed is more or less than the mean historical amount. [3]

Switzerland, like many other European countries, uses multiple different resources for heating residential, as well as industrial housings and facilities. Gas is the second most used resource, accounting for approximately 18% of the distribution of the energy sources (see Figure 1). [4]

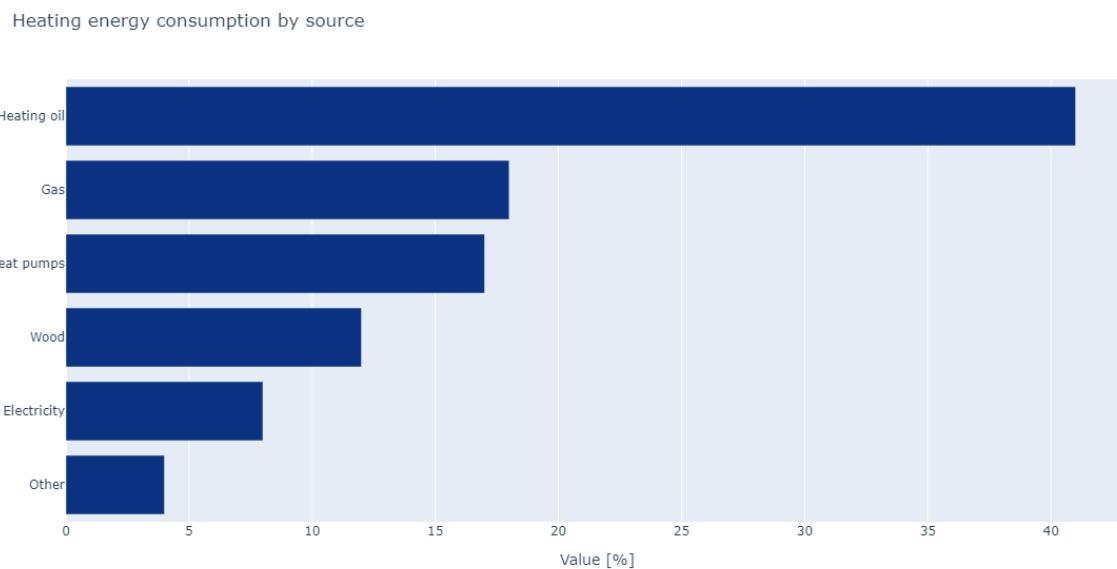


Figure 1 Heating energy by source

In contrast to heating oil, gas is much more susceptible for short term pricing changes which affect the buyers. Oil can be bought and stored in bulk and replenishing the stock can be timed better. Gas on the other hand is streamed form the gas pipelines, with only little reserves stored away. Because of the energy efficiency in comparison to volume, and thus storage space, it is not a common practice to store gas in bulk.

The dependency of Europe and Switzerland on imports of natural gas was made clear during the year 2022 with military aggression of Russia against Ukraine. With 167 billion cubic meters of natural gas imported from Russia in 2021 alone, the country was Europe's major supplier. This number drastically decreased since the beginning of 2022. [5], [6]

With the given circumstances the importance of gas as a resource for Switzerland and the prediction of the needed amount as precisely as possible is apparent.

When forecasting weather, the most important factor influencing the heating demand is the outdoor temperature at a given time. Other factors, which are out of scope for this thesis, are building properties like shape, insulation, age, and parties per apartment, as well as geographical and demographic features. [7]

Reasons for these factors to be out of scope are firstly, that the forecasting is done for Switzerland as a whole, aggregated as one region. Thus, the individual building properties and geographical factors become neglectable as the demand is averaged together. Secondly, these factors would complicate the modelling process as well as the models themselves and in doing so the project would exceed the given time frame.

As previously stated, the long-term forecasting of the weather and temperature is a difficult and complicated topic. An enormous amount of research has already been conducted regarding this field of study (see chapter 2). Different ways of forecasting have thus been explored, ranging from statistical methods, over system-based modelling and machine learning.

The bigger and well-known weather prediction models combine and compile weather measurements from all over the globe to create the forecasts. Such models include, among others, the global forecast system (GFS) and the European centre for medium-range weather forecasts (ECMWF). [8], [9]

Independent researchers often make use of statistical models and machine learning while focusing on certain regions and time spans to create forecasts (see chapter 2.3). This thesis makes use of machine learning as well as more statistical methods. Which models were chosen exactly is dependent on the literature review and the model performances.

A large contributing factor for the model's skill are the input features. There is a big variety to choose from. The big models like ECMWF and GFS make use of as many features and measurements as possible. But this approach is not feasible for smaller and more specific machine learning models, because of the complication they pose. Most often local measurements of atmospheric metrics are taken into account, as reviewed in chapter 2. Further features can be more global weather phenomena which affect certain areas on the globe. What these features are and how they can be used is a big part of the literature review as well as the modelling process.

One of the more global weather phenomena which is explored in more detail, is the polar vortex. To incorporate polar vortex trend data into the model as a feature, multiple different approaches of detecting breakdown- or weak-events were explored, but mostly limited to mathematical and statistical methods. There are some indexes which cover aspects of the polar vortex, such as the North Atlantic oscillation and Arctic oscillation. However, there is no specific approach for determining the impact of certain states of the vortex for the exact region of Switzerland. [10], [11]

## 1.2 Goals and Research Question

The objective of this thesis can be divided into two different parts, which build on one another. As stated in the previous chapter, the overarching goal is to predict the next month's average temperature within Switzerland for the use of energy trading. The two, or three respectively, main components, which make up this project are as follows:

- Outdoor temperature prediction (primary objective):  
The prediction of the outdoor temperature itself plays a crucial role. If some correlation can be found as stated in the previous paragraph, the temperature forecast would gain said insight. Two methods are used, which make up the other two components of this project. A statistical approach (for the polar vortex) and a machine learning approach. Furthermore, multiple

different machine learning models are tested, and the resulting optimized models are compared to each other.

Because this is not a novel idea as explored in chapter 2, a great focus will be set on the data exploration and feature engineering. But generally speaking: The combination of local weather metrics and global weather phenomena to create a more local forecast was not used to a great extent by any of the reviewed independent literature in chapter 2. The mentioned studies mostly focused on either local forecasting with general weather metrics like temperature and pressure, or global scale forecasting by regarding influences like the El Niño-Southern Oscillation phenomenon or the Arctic oscillation. Some models used one of the more global weather phenomena, but not a combination of them, in conjunction with local measures. Furthermore, the polar vortex was only considered by the means of the Arctic oscillation and North Atlantic oscillation index. This particular weather occurrence has a great impact on temperature within Europe and will thus be explored and enriched further to generate a deeper insight into its effects on Switzerland [12], [13]. Additionally, none of the explored areas within the listed papers explored the country of Switzerland.

- Gas price and temperature correlations (secondary objective):  
How could the prediction of the outdoor temperature be linked on forecasting the gas prices. Gas prices are amongst other things influenced by the outdoor temperature, because of its uses in heating [14]. Therefore, if it would be possible to predict the next month's average temperature, some inference could theoretically be made to the gas prices. Thus, allowing for better planning and sourcing of gas needs. Furthermore, it would aid in the financial planning of the energy departments and energy traders. In order to achieve this, a model, equation, or algorithm would be necessary to link the two together. As stated in chapter 2.1, some research has already been done regarding said subject. But an exact description model was not provided.

The research questions can hence be derived as follows:

- What are important features for temperature forecasting in Switzerland and how can the polar vortex be utilized or data on it be transformed to be used as a feature?
- What machine learning models and methods can be used to predict next month's average temperature and how accurate are they?
- How does the outdoor temperature within Switzerland influence the gas prices?

## 2 Literature

### 2.1 Influences on Gas Prices

As stated in an article of the Acta Montanistica Slovaca journal, there are multiple different factors which influence the gas prices. The article considered eleven different OECD (Organisation for Economic Co-operation and Development) countries, among these is Switzerland. The objective was to examine the effects of climate and oil prices on residential natural gas prices. In order to achieve this goal, panel unit root tests were applied (CCEP). The CCEP is a statistical method, which is used to estimate different parameters of a linear regression model [15]. With the used methods, the author of the article was able to proof that, generally speaking, the heating degree days (HDD) have a negative impact on the natural gas prices, consumed in the residential sector. Whereas the HDD is defined as an index which represents the amount of natural gas required for the household sector during the cool season. Although, it has to be mentioned, that the given coefficient for the HDD index for Switzerland is rather low, compared to other countries mentioned in the research. [14]

An article, published in the Energy scientific research paper, created a multi-scale perspective on the behaviour mechanism of regional natural gas prices. Among other factors, short term events which have an impact on the gas price, were explored. Besides the main driving factors (such as oil prices, demand, and supply) it found that changes in temperature during the winter seasons can cause fluctuation on the gas prices (on the Henry Hub spot prices to be precise). Switzerland was not featured explicitly in the work. But there was some information to be found about Germany, which can be used as a reference to some degree. A SVAR model (Structural Vector Autoregression) showed that natural gas prices were influenced by temperature, storage, and supply shortages in the short term. Furthermore, the results yielded that fluctuations in prices are linked to seasonality. [16]

### 2.2 Weather Forecasting Models

Some of the most used weather forecasting models on either a global scale or for Europe specifically include the global forecast system (GFS), the European centre for medium-range weather forecasts (ECMWF), and the icosahedral nonhydrostatic weather and climate model (ICON). [17], [18], [19]

GFS: The GFS is one of the most used weather forecasting systems globally. The model is updated every 6 hours and consists of 4 separate sub-models. Each of these is responsible for one of the following components: atmospheric, ocean, land / soil, and sea ice. It features a maximum of a 16-day forecasting horizon or lead time. [20], [21], [9]

ECWMF: The ECWMF is a mathematical model, with 34 countries involved in its development and maintenance. Despite its name, the forecasting model is not limited to Europe. It provides an accurate forecast for medium lead times (up to 15 days) as well as extended (6 weeks) and long range (4 months) forecasts for temperature, among other metrics. However, the extended forecast (6 weeks) consists of anomalies in temperatures 2 meters above the ground. The long-range model (4 months) only provides probabilities, based on terciles for temperatures. [22], [23], [8]

ICON: The ICON consists of a suit of models which were created and maintained by Germany's national weather service. Three of the major ICON models are ICON-D2 for Central Europe, ICON7 for Europe, and ICON13 for global forecasting. A range of maximum 7.5 days can be reached. [24], [25]

### 2.3 Temperature Forecasting

Applying machine learning in order to forecast weather is not a novel idea. There have been numerous attempts to create accurate weather and temperature forecasts with machine learning.

One such study, published by the journal Energies, compared different approaches and models. It showed that a forecast with input features, such as temperature, relative humidity, solar radiation, rain, and wind speed measurements can yield somewhat accurate results. Furthermore, they compared traditional neural networks to deep neural networks, which performed better on one-step ahead forecasting of regional temperature. Said models yielding a mean squared error of 0.0017 °C. For global scale forecasting, support vector machines were preferred, because they provide a good balance between model complexity and accuracy. [26]

On a closer inspection of the input features of the used models for long term global predictions the El Niño Southern Oscillation, Solar Irradiance, CO<sub>2</sub> measurements, and stratospheric optical depth were also used as input features. Furthermore, monthly temperature forecast models were looked at. All of the different models used regression for the forecasting. The two most used model types were multi-layer perceptrons as well as support vector machines. [26]

An approach for short term forecasting (one day ahead) of weather, developed for Iran, used a multi-layer perceptron model. The utilized input features included: Wind speed, wind direction, air temperature data, relative humidity, dew point, pressure, visibility, and the amount of cloud coverage. It was found that a multi-layer perceptron with one hidden layer and 6 neurons achieved the best outcome. The mean absolute error ranges from 0.0079 to 1.2916 °C. [27]

In an article published in 2003 in the Engineering Applications of Artificial Intelligence journal, short term forecasting of temperatures by using abductive networks was proposed. Said networks provide the advantage of simplified and automated model synthesis, while providing more transparency for input-output models. Performance wise, the model produced a mean absolute error of 0.93 °C for daily and 0.58 °C for weekly predictions. It was stated that the performance is superior to naive forecasts bayes, which is based on persistence and climatology. The used dataset consisted of mean and standard deviation in an hourly resolution from weather stations within the USA. [28]

Attempting to forecast atmospheric temperatures on a mid-term range (2 to 10 days), research was conducted focusing on support vector machines. It found that the tested machine learning models worked better than traditional statistical models. Multiple different kernel functions were tested to achieve the optimal forecast, settling on the radial basis function. In the end, the support vector machine was compared to a multi-layer perceptron, concluding that it performed consistently better than the neural network. The used input features were not listed in detail, but it was mentioned that “weather data” was used. [29]

A paper researching the viability of long-term forecasting of temperatures in Iran, made use of the statistical model seasonal autoregressive integrated moving average, as well as support vector regression with and without the firefly algorithm for optimization. It concluded that the seasonal autoregressive integrated moving average model outperformed the support vector machine, despite its linearity. The root mean square error was 1.027 °C. Between the support vector machines, the one utilizing the firefly algorithm for optimization produced a smaller error, but both are still viable options. Unfortunately, neither the dataset nor the input features are listed in the article. [30]

Long short-term memory cells were also researched in an article for the Communications in Statistics - Simulation and Computation magazine for temperature forecasting. In addition, a prophet model by Facebook was implemented for comparison. The long short-term memory model which was used is a single variant model, only using the historical air temperature as an input feature, transformed in multiple ways. Different combinations of parameters were tested to find the optimal neural network, with the following values coming out on top: 30 epochs of training, 1 batch size, 1 hidden layer, 1 cell in the hidden layer for maximum air temperature, 10 cell for minimum air temperature, adam optimizer. The root mean squared error on predictions of maximum and minimum temperature were 1.23 °C and 0.94 °C respectively. [31]

Another study dealing with deep neural networks, published in the Applied Soft Computing journal, mainly compared three different computational frameworks for the subject of air temperature predictions:

- Two convolutional neural networks. One using video to image translation and the other utilizing recurrence plots, which convert time series to images.
- One model which made use of LASSO regression and decision trees.

9 different, weather-related input features were considered for the models: air temperature, sea surface temperature, 10 m and 100 m u (eastward) and v (northward) wind components, mean sea level pressure, volumetric soil water layer, and geopotential pressure level on 500 hPa. The models themselves predict temperatures for one and two fortnights lead time in France and Spain. Looking at the results of the models (a convolutional neural network with recurrence plots and binary encoding), the best ones for the two-fortnight lead time in France had a mean squared error of 2.117 °C. [32]

Contradicting the previous findings on long-term forecasting with neural networks, an article featured in the scientific journal Water comes to different conclusions. It reviewed, once again, different approaches to temperature forecasting with machine learning. Relevant for this study were the multi-layer perceptron and recurrent neural networks with long short-term memory cells. It was found that the neural networks are mainly viable for short-term predictions of air temperatures. [33]

The root mean squared error of a tested model for average monthly temperature forecast within Turkey ranged from 0.705 to 2.600 K. Unfortunately, the features were not listed for this model. Furthermore, a comparison between two models, a neural network and a support vector regression for the region of Oceania, revealed that the latter one provided more accurate forecasts. The following input features were used: Southern Oscillation Index (SOI), Indian Ocean Dipole (IOD), and Pacific Decadal Oscillation (PDO). [33]

## 2.4 Temperature Anomaly Detection

Apart from forecasting the exact temperature, it is also interesting to further inspect the anomaly forecasting of temperature on different time scales. Although not many papers or methods could be found regarding said topic.

One of the studied sources provided a comparison, looking at twelve competing models for predicting global anomalies in temperature. This ensemble of models contains, among others, an autoregressive integrated moving average (ARIMA), exponential smoothing, and neural networks. The two used features were the global carbon dioxide emissions and global temperature anomaly. The latter one, was a custom indicator, calculated by the difference between a reference long-term average value and the current value. The forecasting itself was done as a regression with multi step forecasting for short, medium, and long term. The overall best model came out to be the HMSSA-V model with an average root mean squared error over all time spans of 12.95 °C. [34]

Another paper looking at the long-term prediction of global temperature anomalies forecasting, proposes the linear inverse model for decadal forecasting. Among other input features like surface and sea level temperatures, as well as moving averages with a lag of a year, the El Niño Southern Oscillation and Pacific Decadal Oscillation are mentioned as influence factors. [35]

Regression is not the only way anomalies in weather data can be detected. By means of a density-based clustering algorithm an article published in Journal of Physics explores the detection of the mentioned anomalies. The model in question makes use of 8 features, such as temperature, humidity, sun exposurer time, windspeeds, and rainfalls. Although, this method was not used to create any sort of prediction, some insight can be gained, by looking at the used input features and their importance in the principal component analysis. Some of the most important ones are the maximum and average temperature, as well as the average humidity. [36]

## 2.5 Weather Phenomena and Effects

There are multiple different weather phenomena, which can impact the outdoor temperature globally, as well as in Europe, in the long term. Some of these, which were previously mentioned and used in machine learning and statistical models, are: (see chapters 2.2, 2.3, 2.4):

- El Niño-Southern Oscillation (ENSO) [37]
- Madden-Julian Oscillation (MJO) [38]
- Polar Vortex (PV) [39]
- Arctic Oscillation (AO) [10]
- North Atlantic Oscillation (NAO) [11]

The ENSO is one of the best explored and well-known weather phenomena. It describes the warm and cool phases, as well as changes in pressure and rainfall of a recurring weather pattern, taking place across the tropical Pacific. It is split up into the two different phases, El Niño and La Niña (see Figure 2). Typically, a shift from one phase to the other happens in an interval of two to seven years. This leads to forecastable changes in ocean surface temperature, wind as well as windspeeds, and rainfall. Said changes can lead to global side effects, impacting temperature and atmospheric pressure, among other factors. [37]

A paper published by the ECMWF explores the impact of the ENSO on European weather and temperature. It states that there are at least two pathways by which the tropical pacific, linked to the ENSO, has an impact on the Atlantic-European seasonal climate. To be precise, during an El Niño phase a negative winter pattern in the NAO and colder winter anomalies over northern Europe can be spotted. [40]

One ENSO indicator is the Southern Oscillation Index (SOI). It is calculated by differences in pressure between Tahiti and Darwin. A negative SOI refers to the El Niño and a positive SOI to the La Niña phases. [41]

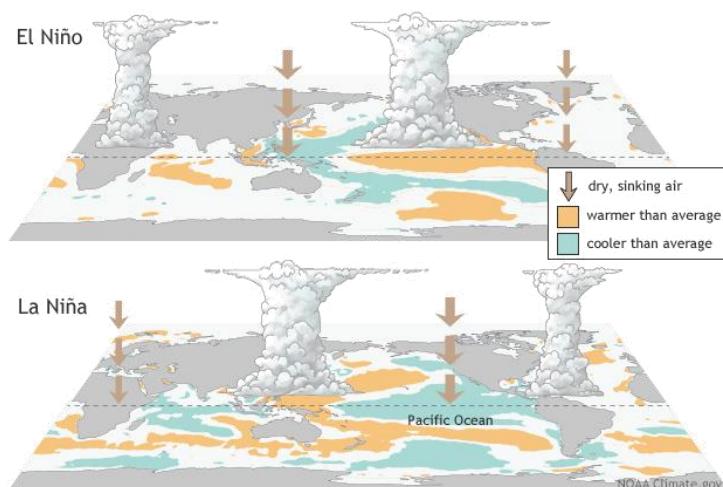


Figure 2 ENSO [37]

The Madden Julian Oscillation index (MJO), in contrast to the stationary and switching ENSO, is an east moving disturbance of pressure, winds, and rainfalls. After traversing along the equator for 30 to 60 days on average, it returns to the starting position. It is possible for multiple MJO events to occur during a season. Hence, it can be described as an intrapersonal tropical climate variability. The progression of the MJO is split up into 8 phases, describing its eastward progression and effects on different regions. As it is the case with the ENSO, the MJO's effect on weather and temperature are well documented. It is important for extended weather forecasting for North America and Europe, due to the changes in windspeeds and rainfall which effect the tropics and extra-tropics. [38]

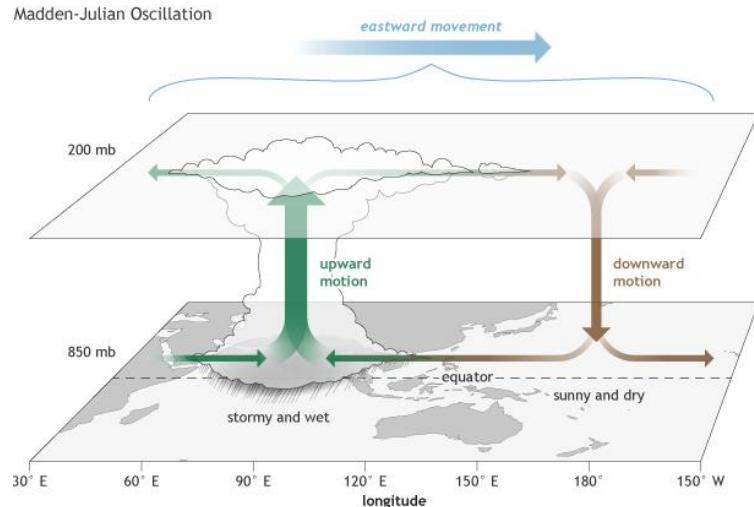


Figure 3 MJO [38]

The before mentioned effects of the MJO are numerous. Two noteworthy ones are: Its impact on the North Atlantic weather regime, especially the North Atlantic Oscillation (NAO) and the statistically significant influence on the predictability on the extra-tropic on sub-seasonal time scales. [42]

Representing the MJO with numerical values requires three parameters, while a fourth can be derived from the first two. Due to the fact that it is a moving disturbance, values are needed correlating to its position, as well as its intensity. This is done by providing a coordinate system with two axes, RMM1 and RMM2. The distance from the centre (amplitude or length of the resulting vector) representing the strength of the MJO. The grid is split up into the previously mentioned 8 phases, each referring to a geological location. [43]

The polar vortex is a circular wind, occurring over the north pole in the stratosphere at altitudes ranging from approximately 15 to 50 kilometres. Although the wind speeds are varying, depending on the current season, the vortex is always present. A weaker or stronger polar vortex comes with respective changes for the northern hemisphere. During the winter months, the polar vortex is typically stronger and therefore more stable in nature. This prevents the cold polar air from drifting further south. When the vortex slows down or weakens, the cooler air can escape southward, bringing cooler temperatures to Northern Europe and North America. But in addition, it also draws in the warmer jet streams, leading to south winds in certain regions. [39], [13]

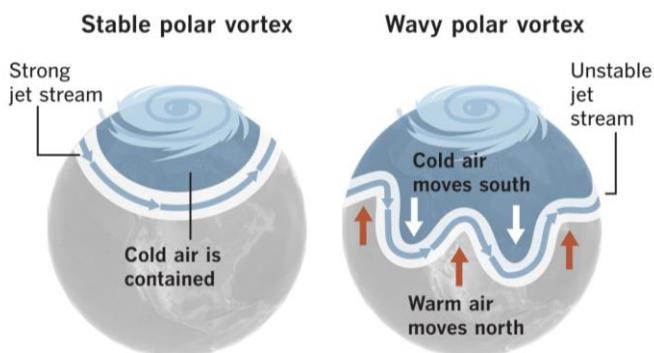


Figure 4 Polar vortex [12]

Similar to the SOI and MJO, the polar vortex can be described numerically by the AO to some degree. It is constructed by projecting height anomalies between 20° north and the north pole onto a typical pattern of the AO. The height anomalies are calculated by taking measurements of air pressure at the

1000 hPa level (near earth's surface) and comparing the altitude values to historical averages of pressure at said height. [10]

Although the AO was not mentioned or used explicitly in any of the reviewed literature in chapters 2.2, 2.3, and 2.4, it has to be considered, due to it being impacted by the ENSO and MJO, as well as the AO's influence on European weather.

The NAO is closely related to the AO. The NAO is calculated by the difference in sea-level pressure between the subpolar low and subtropical high. A positive value correlates to below normal heights and pressure, and a negative value above normal heights and pressure at higher latitudes of the North Atlantic Ocean, Western Europe, and Eastern United States. Changes of the NAO inflict changes on the North Atlantic jet streams and on changes in temperature for the before mentioned regions. [11]

In order to analyse, incorporate, and properly assess the polar vortex, a comprehensive list of breakdown or weak events would be needed. But a suitable list did not seem to be available as an online resource. Because of this, a compendium of sudden stratospheric warming events (SSW) was consolidated as a guideline (see chapter 3.5). This weather phenomenon describes the sudden rise in air temperature in the stratosphere above the north pole. Within a week the temperatures can increase by 50 K. One of the effects is, that the polar vortex often gets pushed away from the north pole. It is also possible that such an event can cause a breakdown of the polar vortex, but it is not a guarantee. [44], [45]

Numerous similar weather phenomena and indexes to describe them do exist and were also mentioned in some papers on temperature forecasting, explored in previous chapters, but will not be further explored. Reason being the increasing complexity with more dimensionality in the data and the given time constraints. These phenomena include, but are not limited to:

- Indian Ocean Dipole (IOD)
- Pacific Decadal Oscillation (PDO)

## 2.6 Literature Summary

As described in chapter 2.1, the influence of the outdoor temperature on the gas prices has been proven to exist. The exact relation to the average temperature over a certain time span has yet to be calculated.

The three explored weather forecasting models in chapter 2.2 provide some information about temperature, but as the sources state, they are mainly powered by statistics and mathematical models. Furthermore, the forecasting of the temperature for longer ranges is most often an anomaly indication, rather than an exact value.

With the considered literature on the topic of temperature forecasting in chapter 2.3 and 2.4 it can be said that there were previous attempts to make short- and long-term temperature forecasts. Most of the found resources used different algorithms, statistical methods, and machine learning models to analyse and select data, as well as creating and optimizing the models. The previously used and explored methods can be summed up as follows:

- Statistical models:
  - o Autoregressive integrated moving average (ARIMA)
  - o Seasonal autoregressive integrated moving average (SARIMA)
  - o Structural Vector Autoregression (SVAR)
  - o Least absolute shrinkage and selection operator regression (LASSO)
- Machine learning models:
  - o Support vector regression (SVR)
  - o Facebook's prophet model (FPM)
  - o Abductive networks (AN)
  - o Multi-layer perceptron (MLP)
  - o Recurrent neural network (RNN)

- Convolutional neural network (CNN)
- Algorithms and various:
  - Particle Swarm Optimisation (PSO)
  - Firefly Algorithm (FFA)

When it comes to input feature, and the decision on building a single- or multi-varient model, the latter ones seem to be more widely used among the reviewed literature. Some of the most used attributes can be split up into the following two groups.

- Weather data:
  - Temperature
  - Wind speeds
  - Humidity
  - Atmospheric pressure
  - Solar irradiance
- Atmospheric phenomena:
  - El Niño-Southern Oscillation (ENSO)
  - Madden-Julian Oscillation (MJO)
  - Polar vortex (PV)
  - Indian Ocean Dipole (IOD)
  - Pacific Decadal Oscillation (PDO)

There are multiple different weather occurrences which have a global impact on weather and temperature. Some of these were also used in the looked at research papers to predict temperatures:

- El Niño-Southern Oscillation (ENSO) / Southern Oscillation Index (SOI)
- Madden-Julian Oscillation (MJO)
- Polar vortex (PV) / Arctic Oscillation Index (SO)
- North Atlantic Oscillation (NAO)
- Indian Ocean Dipole (IOD)
- Pacific Decadal Oscillation (PDO)

## 3 Methods

### 3.1 Tools

The accompanying code was written in python in the file format of jupyter notebook (.ipynb). A local runtime of python, version 3.10.8, was used. The programming environment was Visual Studio Code. The nonstandard libraries used during the project were: pandas, numpy, plotly.express, sklearn, tensorflow, keras, clustimage.

### 3.2 Research Design Overview (CRISP-DM)

The throughline of the methods is based on the Cross Industry Standard Process for Data Mining (CRISP-DM), which is usually applied in the data science community. It consists of the depicted steps, shown in Figure 5. The last step “Deployment” is out of scope. The detailed application of each step is listed in the following chapters. The general description is listed below. [46]

This approach was only applied for the temperature prediction, but not the linkage of gas prices to outdoor temperature and the image clustering of the polar vortex data.

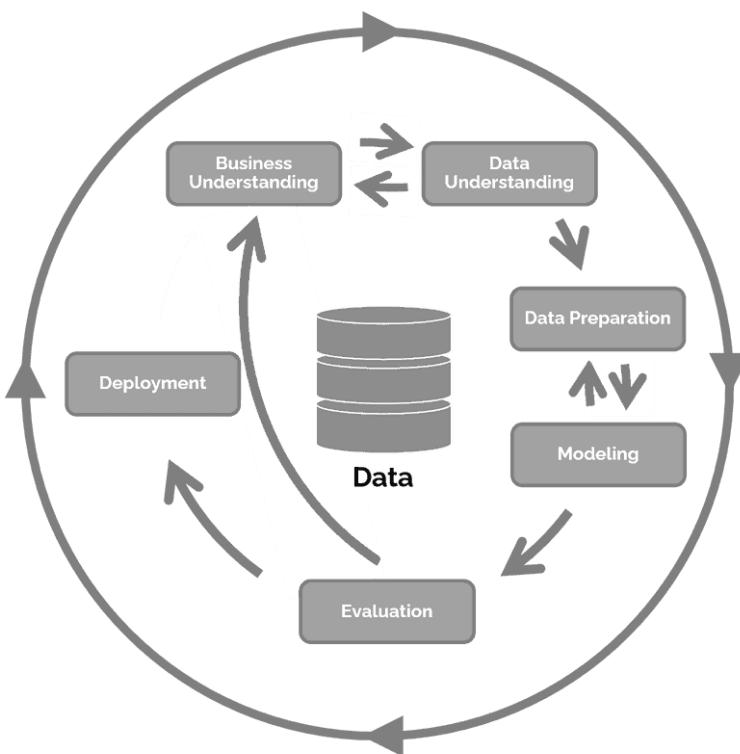


Figure 5 CRISP-DM [46]

**Business understanding:** The first stage consists of defining the business objective relevant for the project. A foundational knowledge of the topic at hand is needed in order to do so. Additionally, this step includes assessing the situation, setting a goal for the data mining or gathering, and creating a general project plan.

**Data understanding:** After the circumstances and the project goals are set, the initial data collecting needs to be conducted. Next, an understanding of the gathered data is needed to proceed, mostly done by data exploration. This can include the analysis of numerical values as well as plotting data.

Data preparation:	In order to use the data as input features for the models, the data needs to be cleaned, merged (if there are multiple data sets), and reformatted. This can include the addition of new features based on the existing data, also known as feature engineering.
Modelling:	The modelling process itself includes the splitting of the data and the model selection, based on the premise and what is most the suitable method to solve the given problem. Normally, multiple different models are created and compared to one another.
Evaluation:	Finally, the model outputs or results get evaluated by previously defined metrics. Furthermore, the follow-up steps are determined.

### 3.3 Business Understanding

The core subject matter consists of two parts. On one hand the correlation of gas prices to the outdoor temperature within Switzerland, and on the other hand the topic of weather and temperature forecasting. With the given literature research in chapter 2 a base understanding has been built up for both listed subjects. Furthermore, the object was also stated in the introduction in chapter 1.

### 3.4 Data Understanding

First, the needed data was gathered. The chosen input feature consisted of the ones listed below. The reason for choosing is stated below (sourced from the respectively listed website):

General weather data [47]:

- Most of the explored literature and especially the machine learning models used at least general weather data as measurements for multi-varied models. Commonly used features among different models, which were used from here on out, were: the outdoor temperature 2 meters above ground (t2m), u and v components of the wind speed as common vector notation, mean surface level pressure, and the clear sky solar irradiation.
- The chosen area for the before mentioned features is given by the below listed geo borders. Note, that the border has been simplified. Otherwise, each obtained data point from the used API would have to be checked against a defined border. This would be a very time-consuming process, with only a minor benefit:
  - o North (latitude): 47.8
  - o East (longitude): 10.5
  - o South (latitude): 45.8
  - o West (longitude): 6.0

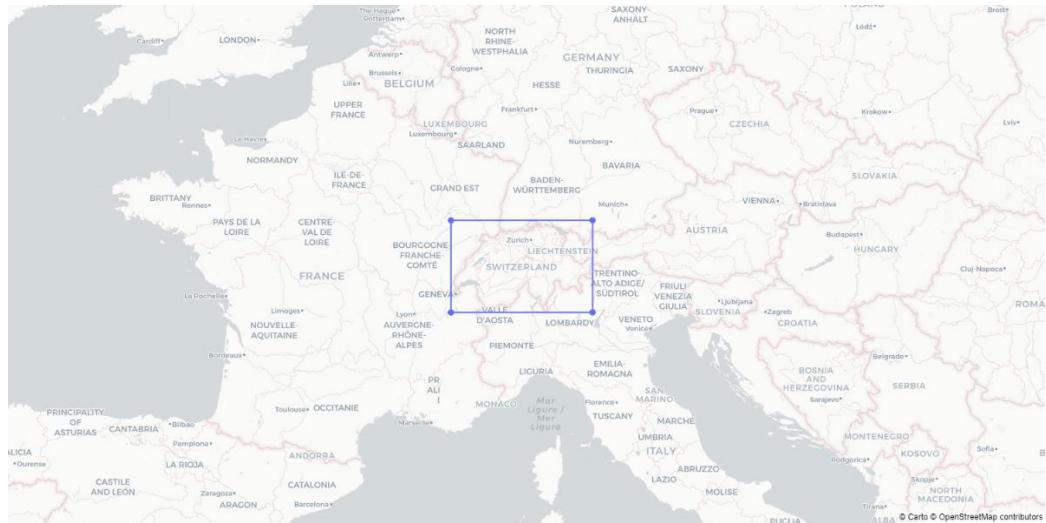


Figure 6 General weather data area map

- The obtained features were aggregated with the mean, firstly over the entire area, then by day. The aggregation over the area was done because the forecasting of the temperature is done for Switzerland as a whole, not split up into different areas.

$$\bar{x} = \frac{\sum x}{N} \quad 1 [48]$$

#### SOI [47]

- The SOI was chosen, because of its documented effects on weather, as explored in chapter 2.5. Furthermore, within the literature review, more than one model also made use of the ENSO in some form for long time weather predictions.
- During the data gathering, no source could be found which provided the SOI in a daily resolution, covering all years ranging from 1979 until 2022. The reason is that the index is normally calculated on a monthly base. Thus, it had to be created manually and compared to an online source for validation. The SOI is calculated as the standardized delta in pressure difference between Darwin, Australia and Tahiti. The coordinates relevant for the calculation were:
  - o Darwin: latitude: -12.46 / longitude: 130.84
  - o Tahiti: latitude: -17.54 / longitude: -149.56
- As previously stated, the SOI is usually calculated on a monthly basis, hence the formula was slightly altered. Instead of using the long term monthly mean and standard deviation of each month, the SOI was calculated for each day and was inverted to match to the northern hemisphere.

$$SOI = -1 * \frac{(Pdiff - Pdiffav)}{SD(Pdiff)} \quad 2 [49]$$

*Pdiff: Mean sea level pressure Tahiti (daily) – Mean sea level pressure Darwin (daily)*

*Pdiffav: long term average of Pdiff for the day in question*

*SD(Pdiff): long term standard deviation of Pdiff for the day in question*

$$SD = \sqrt{\frac{\sum |x - \mu|^2}{N}} \quad 3 [50]$$

## MJO [51]

- The MJO is similar in impact on weather forecasting as the SOI, as explored in chapter 2.5 and is also used for multiple different models mentioned in the literature review. Because the data source provided the MJO in a daily time resolution, no transformation was needed. The missing values were dropped, which resulted in a useable and continuous range of values from 1979 to 2022.

## AO and NAO [52], [11]

- The polar vortex measured in the form of AO and NAO has, as documented a great impact on the temperature within Europe. Furthermore, the indexes are closely related to one another. The AO and NAO are two ways of representing the polar vortex. Once again, the data is provided in a daily resolution, thus no aggregation was needed. The missing values were filled using a built-in function of the pandas library called linear interpolation. The linear interpolation is defined as:

$$y = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad 4 [53]$$

## Custom polar vortex data [54]

- Because the AO and NAO are calculated quite simplistically, a more detailed look into the polar vortex was needed. It is one of the main influences on long-term weather for Europe [39]. The raw data for this assessment consists of the u and v wind components and the temperature at the pressure levels of 100 to 10 hPa (Note that not all steps were provided by the used API). These pressure levels roughly represent the altitudes of 15 up to 30 km [55], [56]. Other attributes could also be considered, but the resulting amount of data is already quite large. To analyse and process even more features would have exceeded the set time frame.
- The before mentioned features were taken from an area defined by the following coordinates. This area spreads from Switzerland up to the north pole. With this approach, a simplified version of the impact of the polar vortex can be analysed and later be broken down into a custom indicator and used for clustering.
  - o North (latitude): 90.0
  - o East (longitude): 9.0
  - o South (latitude): 45.0
  - o West (longitude): 8.0
- Because it would be too much data to handle in a reasonable amount of time, the geographical resolution was set to one data point for each whole longitude and latitude grid and pressure level. All datapoints were averaged together with the mean, first over the defined grid, and afterwards to a daily time resolution. [48]
- Because of the data frame's size and therefore the performance to calculate the data, some feature engineering was already done during this stage. Added to the data frame were two new columns: wind speed and wind direction. Each of these features was calculated for each grid point and time step.

$$\text{wind speed [m/s]} = \|\vec{v}\| = \sqrt{u^2 + v^2} \quad 5 [57]$$

$$\text{wind direction [deg]} = \tan^{-1} \left( \frac{v}{u} \right) * \frac{180}{2\pi} \quad 6 [57]$$

After gathering the data, it was explored by numerical means as well as with different plot types. One of the numerical explorations is done with the correlation matrix, based on the Pearson correlation coefficient. It describes the linear correlation between two values or arrays [58]:

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad 7 [58]$$

### 3.5 Data Preparation

The data preparation was split up into three parts: the feature engineering of the main data frame; the exploration and development of a custom polar vortex index; the data joining and preparation for the machine learning model, excluding the standardizing, data splitting and shifting window implementation.

It is usual to transform datetime like objects or variables by the means of cyclic functions. The machine learning models are not able to understand the cyclic nature of years or days, only by a constantly increasing value. Thus, the sine and cosine function were applied on a yearly scale. Meaning that one cycle of the functions represents the passing of one year. Both have to be used in conjunction, because if only one is provided, there is the same value, representing two points in time, as the functions oscillate between -1 and 1. If both are used, this ambiguity can be prevented, and each combination of the resulting values form the two functions represent a unique point in time on the defined yearly scale. [59]

$$\text{year sine} = \sin\left(t * \frac{2\pi}{n}\right) \quad 8 [59]$$

$$\text{year cos} = \cos\left(t * \frac{2\pi}{n}\right) \quad 9 [59]$$

*t: timestamp*

*n: number of seconds in a year (31'622'400)*

To represent the idea of the current trend in the data, a moving average (rolling mean) was applied to the variables AO, NAO, SOI, and t2m. Like the before done transformation, this is common practice when working with time series or time related data and predictions. The impact of these variables cannot yet be determined because the machine learning models use shifting windows to feed in past data points. Generally speaking, if the current value is bigger than its moving average, the metric is an uptrend and vice versa. The value of n was set to 30 days. [60]

$$ma = \frac{x_t + x_{t-1} + \dots + x_{t-n}}{n} \quad 10 [60]$$

Next, the wind speed and direction were added in the same manner as done with the polar vortex data by applying the respective formulas (see page 18). The variables used were the u10 and v10 component of the wind speeds.

In order to create a more continuous spectrum of data across the pressure levels for the polar vortex data, all features were synthesized and recalculated for all pressure levels. The concerning levels were: 40,60,80,90 hPa. This provides a more complete picture with one step every 10 hPa. Firstly, the data was interpolated linearly and merged into the main pv data frame. Secondly, the wind speed and direction were calculated from interpolated wind speeds. For the sake of simplicity, the gained altitude for the increasing pressure levels in 10 hPa steps is assumed to be linear.

$$y_{ip} = y_1 + (x - x_1) \frac{y_2 - y_1}{x_2 - x_1} \quad 11 [61]$$

Four custom indexes with different variations were developed and tested to create a simple version to represent a breakdown event of the polar vortex. Each of the indexes was calculated based on the daily data, so no historical data from the set time point were used. Except when trying to detect breakdown events on longer term time scales.

**Index 1:** Index 1 was based on the consistency of wind speeds across multiple pressure levels. The idea was to locate the vortex by identifying the latitude with the smallest deviation in wind speeds, which simultaneously has a high overall wind speed. The applied algorithm uses the formula below to get a value for each latitude, then the latitude with the maximum value for that day was chosen as the value of the index. Important to note is, the wind speed standard deviation is normalized across each day to get a “weight” value between 0 and 1 [62]. The inversion is done to weigh the wind speeds with the lowest deviation the highest.

$$lat_{i1} = ws_{mean} * \left( 1 - \left( \frac{ws_{std} - ws_{min\ std}}{ws_{max\ std} - ws_{min\ std}} \right) \right) \quad 12$$

*ws<sub>mean</sub>: mean latitude wind speed across all levels*

*ws<sub>std</sub>: standard deviation of the wind speed at the set latitude across all pressure levels*

The second variation of the index also incorporated the temperature. The base weight was once again the inverted and normalized standard deviation of the wind speed to keep the same basic behaviour. The mean temperature was included in the same way. Because the lower temperatures give an idea on where the border of the polar vortex is currently located. Both of these can be weighted manually, but the sum of both should not exceed 1.

$$lat_{i1} = ws_{mean} * \left( w_{ws} * \left( 1 - \left( \frac{ws_{std} - ws_{min\ std}}{ws_{max\ std} - ws_{min\ std}} \right) \right) + w_t * \left( 1 - \left( \frac{t_{std} - t_{min\ std}}{t_{max\ std} - t_{min\ std}} \right) \right) \right) \quad 13$$

*ws<sub>mean</sub>: mean latitude wind speed across all levels*

*ws<sub>std</sub>: standard deviation of the wind speed at the set latitude across all pressure levels*

*w<sub>ws</sub>: weight wind speed*

*t<sub>mean</sub>: mean latitude temperature across all levels*

*t<sub>std</sub>: standard deviation of the temperature at the set latitude across all pressure levels*

*t<sub>ws</sub>: weight temperature*

**Index 2:** The second index is based on sharp increases or decreases (also called edges) in wind speeds or temperatures across all given latitudes. Reason being, that a sharp edge, especially in wind speeds can mark the border of a stable and strong polar vortex. The algorithm which calculated the index also allowed to set a threshold which has to be exceeded before an edge is identified. The number of latitudes for the comparison of increase or decrease can also be chosen freely.

$$lat_{i2} = \begin{cases} \max[d_1, d_2, \dots, d_n] \geq d_{mean} * th, & lat_{d\ max} \\ \max[d_1, d_2, \dots, d_n] < d_{mean} * th, & 0 \end{cases} \quad 14$$

$$d = abs(m_1 - m_2) + abs(m_2 - m_3) + \dots + abs(m_{n-1} - m_n)$$

*m: the chosen metric, either wind speed or temperature*

*th: the chosen threshold as decimal*

Due to the fact, that this method allows there to be no edge or latitude as value for the index, a breakdown event can be defined and automatically marked as such. Two additional parameters were programmed into the class to allow optimization: the

breakdown offset, which defines the number of past values to be considered; the breakdown sensitivity, which defines at which threshold a breakdown is detected. A rolling mean was applied to the `pv_edge` attribute over the last n periods, which describes if there is an edge or not as a boolean in integer form (meaning 0 or 1). Whereas n is equal to the breakdown offset. If the rolling mean falls under the defined sensitivity threshold, the entry is marked as a breakdown event. This method should prevent detecting breakdown events, if only a few observations out of the given sample do not show a vortex edge.

**Index 3:** This idea of this index was derived from the method of calculating the SOI. It takes two points at each end of the latitude spectrum, 45° and 89° north, gets a certain metric on each and is represented by the delta of the two points. In order to include all available data, each day was aggregated over the pressure level. A weight between 0 and 1 for each latitude was set, based on the distance for each end of the spectrum. Then the normalized metric was multiplied by said weight, once for each point as well as the delta was calculated. This delta represented the index 3.

$$ind_3 = \left( 1 - \frac{lat - lat_{min}}{lat_{max} - lat_{min}} \right) \left( \frac{m - m_{min}}{m_{max} - m_{min}} \right) - \left( \frac{lat - lat_{min}}{lat_{max} - lat_{min}} \right) \left( \frac{m - m_{min}}{m_{max} - m_{min}} \right) \quad 15 [62]$$

*m: the chosen metric, either wind speed or temperature*

**Index 4:** The last method was the simplest. It consisted of aggregating the values of a chosen metric across pressure levels. The programmed function allowed to set the aggregation to either sum, minimum, maximum, median, or average. With some brute force testing the mean aggregation method was chosen. Afterwards, the method sets the latitude of the maximum (for wind speeds) or minimum (for temperature) aggregated value as the index. With the gained insight from the index 2, the same pattern of detecting breakdown events was then applied. A threshold can be set, which, when exceeded, will mark the existence of a strong polar vortex. The detection of breakdown events was then applied in the same way.

$$lat_{i4} = \begin{cases} \max[m_1, m_2, \dots, m_n] \geq m_{mean} * th, & lat_{m\ max} \\ \max[m_1, m_2, \dots, m_n] < m_{mean} * th, & 0 \end{cases} \quad 16$$

*m<sub>n</sub>: the mean of the chosen metric across one latitude*

*m<sub>mean</sub>: the mean of the chosen metric across all pressure levels and latitudes*

*th: the chosen threshold as decimal*

Each index was compared on some sample data and the most suitable ones were identified by plotting them over the given cluster of latitudes, pressure levels and the used metric (wind speed or temperature). Furthermore, they were compared to the SSW events. The best one for each metric was then further developed and optimized. Finally, each of the indexes was added to the main data frame.

Because the focus is on long term forecasting, the target variables will be two bi-weekly-periods:

- t1: Rolling mean of the upcoming 14 days ( $t2m_t$  to  $t2m_{t+14}$ )
- t2: Rolling mean of 14 days, starting at day 14 until day 28 ( $t2m_{t+14}$  to  $t2m_{t+28}$ )

The split of the monthly period allows on one hand for a well enough aggregation of the data to create rougher estimates for the time periods. On the other hand, it is more detailed than just one target variable, encompassing 30 days. Another advantage is, that the performance on the two periods can be compared to one another, to gain insight in change in accuracy model parameters.

### 3.6 Modelling - Temperature Forecasting

Before the models could be trained, the following three steps were implemented in the process:

- Splitting the data frame into three sets (training, 70%; validation, 20 %; testing, 10%), based on the categorical distribution, while maintaining the temporal order as well as possible. If the data would be shuffled randomly, contextual information would be lost.
- Applying a shifting window of 30 days, to provide past information to the machine learning models. The time frame of 30 was set to match the forecasting period.
- Data standardization to prevent a bias in features, based on their value ranges. This is crucial when working with neural networks. The standardization for the validation and testing set was performed using the mean and standard deviation of the training set, to prevent a bias within the data frames.

The formula applied for standardization the data frames is defined as:

$$z = \frac{x - \mu_{train}}{\sigma_{train}} \quad 17 [63]$$

$\mu_{train}$ : Mean of the training set

$\sigma_{train}$ : Standard deviation of the training set

The model selection was based on the models found in the literature and the most suitable for time series forecasting. The exact reasoning is given below. Note that a detailed explanation of the models, the used learning algorithm and activation functions are not given. Please consider external sources.

#### **Random Forest (RF):**

The RF is an ensemble model, which consists of multiple estimators or trees. The result of the model is decided by averaging the outcome from each tree. A benefit of this model is the insight which can be gained on the feature importance of each given metric. It can be applied to many different use-cases and possesses as a simple base line model, to measure the rest against. [64]

#### **Support Vector Machine (SVM):**

The SVM is a model which creates predictions based on spatial separation. One big advantage of SVM is its high effectiveness on higher dimensional data. Meaning, that model will also perform well, if numerous input features are provided. This was the case, because the input data was windowed, creating more dimensionality. Furthermore, like the RF, it is a very versatile model. [65]

#### **Multi-Layer Perceptron (MLP):**

The MLP is one of the simpler feed forward neural networks, consisting of one input, one or more hidden and one output layer. On each hidden layer there can be an arbitrary number of nodes or cells. It can handle large data well, works on predictions with nonlinear and complex correlations and is highly adjustable. Additionally, through the backpropagation, the network adjusts its weights during the training phase to identify important patterns which impact the output positively. [66], [67]

#### **Recurrent Neural Network (RNN):**

The RNN builds on the basics of the MLP but expands on it by providing long short-term memory cells or nodes. These cells have an internal memory state and can retain information about the previous data points. When working with time series data, this can pose a big advantage, especially for detecting trends and factoring in previous changes in the target value. [68]

Both, the MLP and the RNN, make use of an activation function. These define how the input into a node is transformed before being passed to the next layer. It is one of the most influential parameters of neural networks. Because of the time constraints, this parameter was fixed and not optimized. [69]

Making use of the Rectified Linear Unit (ReLU) activation function for the MLP, fast computation / training posed one of the big advantages, because of its simplicity. Additionally, a fast gradient descent can take place. A disadvantage of the ReLU is, that it is not capped at a certain value and therefore the

outputs are not normalized. Furthermore, it is also possible for it to end up in “dead” state, meaning that high input parameters can cause a shift of the function, which results in only producing zero value outputs as a consequence. The ReLU function is defined as follows: [69]

$$f(x) = \begin{cases} x < 0, 0 \\ x \geq 0, x \end{cases} \quad 18 \text{ [69]}$$

The RNN used the SeLU activation function. This altered version of the ReLU negates the before mentioned problems, while still having access to its advantages. [70]

Due to the fact that the library used to create the MLP does not provide the SeLU activation function, the ReLU was chosen, even though of its disadvantages. The SeLU function is defined as:

$$f(x) = \begin{cases} x < 0, a(e^x - 1) \\ x \geq 0, x \end{cases} \quad 19 \text{ [71]}$$

### 3.7 Modelling – Polar Vortex Clustering

As a further addition to the polar vortex, as well as to the custom indexes, and to propose a simple concept of the further steps, an image clustering of the gathered and pre-processed polar vortex data was created. This was meant as a small proof of concept for the application of the image clustering by means of a k-means model, the process was kept simple and short. This process is, in contrast to the temperature prediction, an unsupervised approach to group the different situation of the polar vortex.

The goal of this process was to cluster the different conditions of the polar vortex. Such datapoints, in form of polar vortex class or label, could potentially be used to feed into another machine learning model which predicts the temperature or other weather metrics.

In order to apply the model, the data was restructured from a grid form to a grey scale image or array, which was standardized [63]. For the clustering no meta data was used, to simplify matters. Hence the set could be split randomly into a training and testing set.

#### K-Means Clustering (KMC)

The K-Means is a clustering method based on centroid algorithm, where the Euclidean distance between each data point and a centroid, which must be assigned, is calculated. With this distance, it is given a suitable cluster. K defines the number of clusters which will emerge. The model was chosen because of its uncomplicated nature and accuracy. Furthermore, it does handle bigger datasets efficiently, and in doing so it reduces calculation times. [72]

### 3.8 Evaluation

The following methods were applied to assess the skill of the respective models and make them comparable. All of the below listed metrics were calculated for t1, t2 and the combined target vector. To get the best general model, the one with the best overall accuracy is set as the top one. The respective metrics are:

#### Confusion Matrix (classification):

The confusion matrix allows to gain a high amount of insight into the model’s performance, when assessing a binary classification. It depicts the number of predicted labels against the actual labels in a two-by-two grid. The used library (scikit learn) provides the confusion matrix in a slightly altered form by inverting the axis. [73]

		ACTUAL	
		Positive	Negative
PREDICTED	Positive	True Positive (TP)	False Positive (FP)
	Negative	5	10
		False Negative (FN)	True Negative (TN)
		15	70

Figure 7 Confusion matrix schema [73]

#### Accuracy (classification):

The accuracy is one of the most used metrics when assessing a classification model. It should only be utilized on a balanced data set. Otherwise, it could lead to false conclusions. It depicts how many values were correctly classified overall, as a relative value ranging from 0 to 1. The goal was exceeding an accuracy of 0.5 for the t1 and t2, and 0.25 for the overall accuracy and therefore providing a better forecast than a random guess. The formula is defined as: [73]

$$acc = \frac{TP+TN}{TP+FP+TN+FN} \quad 20 \text{ [73]}$$

#### R-squared (regression):

The R-squared ( $R^2$ ) error, also known as the coefficient of determination, explains how well the model fits the data. 1 means that a model flawlessly predicts all the variation in the data, and a 0 indicates, that the model is not able to explain any of the variation in the data. Thus, a higher value of  $R^2$  is desired. The formula is defined as: [74]

$$R^2 = \frac{\sum(y_i - \bar{y})^2}{\sum(y_i - \hat{y})^2} \quad 21 \text{ [74]}$$

#### Root Mean Squared Error (regression):

The root mean squared error (RMSE) measures the root of the average mean difference between the predicted value and the actual value. Hence, the lower the value, the more precise the model is. The formula is defined as: [74]

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad 22 \text{ [74]}$$

### 3.9 Gas Price and Outdoor Temperature Correlations

To assess the correlations to the outdoor temperature, the same data was utilized, which was used for the temperature classification models. Additionally, the Natural Gas Spot price in USD was sourced and used as a base for the price. Before analysing the data, all values from both arrays (temperature, natural gas spot prices) were standardized [63] to make them comparable on a visual basis.

The correlations were compared with the following methods and indicators:

- Visualization and visual analysis:  
In order to gain an understanding, the data was plotted in different ways to visually detect correlations between the two factors and detect anomalies.
- The Person correlation coefficient: [58]  
To detect if a linear correlation between the two exists, the correlation coefficient was calculated, which could give an indication about their relationship.
- Robust peak detection algorithm (z-score): [75]  
This detection algorithm is based on a moving average and standard deviation, which are multiplied by a threshold, against which the current value is compared. A python implementation can be found in the appendix (chapter 11.5). This method allows a comparison of the current value to the current trend in the data, which could hint at outliers.
- Manual threshold:  
A manual threshold was implemented as an alternative to the robust peak detection. This method compares each value against a given threshold, after the data was standardized. Hence providing an easy way to detect values which deviate from the norm.

As a comparison, the concluded findings were compared to the respective literature, cited in chapter 2.1.

## 4 Results

### 4.1 Data Understanding – Data Gathering

The general weather data consists of the below mentioned features. These were downloaded from the Copernicus API via an automated script. The time range of the data was set from 01.01.1979 to 31.12.2022. The concerning area was defined as stated in chapter 3.3 and is depicted in Figure 6 in the appendix. To represent the needed features, defined in chapter 3.4, the following attributes were requested from the API and aggregated, by calculating the mean. No data cleaning was needed because no empty values were provided by the API.

API variable	Feature	Unit	Description
2m_temperature	t2m	k	The surface temperature 2 meters above ground
10m_u_component_of_wind	u10	m/s	The eastward component of the wind speed 10 meters above ground
10m_v_component_of_wind	v10	m/s	The northward component of the wind speed 10 meters above ground
clear_sky_direct_solar_radiation_at_surface	cdir	J/m <sup>2</sup>	The net amount of solar (shortwave) radiation reaching the surface of the Earth (both direct and diffuse), under the assumption of cloudless conditions
surface_pressure	sp	Pa	The atmospheric pressure at ground level

Table 1 General weather data - API references

For calculating the SOI, a similar script was written to fetch the data from the API for the previously defined locations. The parameters used for the calculation of the SOI is the mean\_sea\_level\_pressure in hPa. Because the initial data was provided in an hourly resolution, all datapoints without any aggregation were taken for the calculation for the daily mean and standard deviation of the pressure. Following this, the SOI was calculated with the altered formula [49].

In order to validate the calculated index, a second data source [76] was used, which provides the monthly SOI. A temporary data frame was created with monthly values, calculated form the daily index. Apart from a different scaling of the values, they seemed to match up. Thus, leading to the conclusion that the calculated SOI is correct. The difference in scaling most likely originates from the different time resolutions. A monthly and therefore longer time interval and more values for each data point, lead the data to converge to a certain value range.

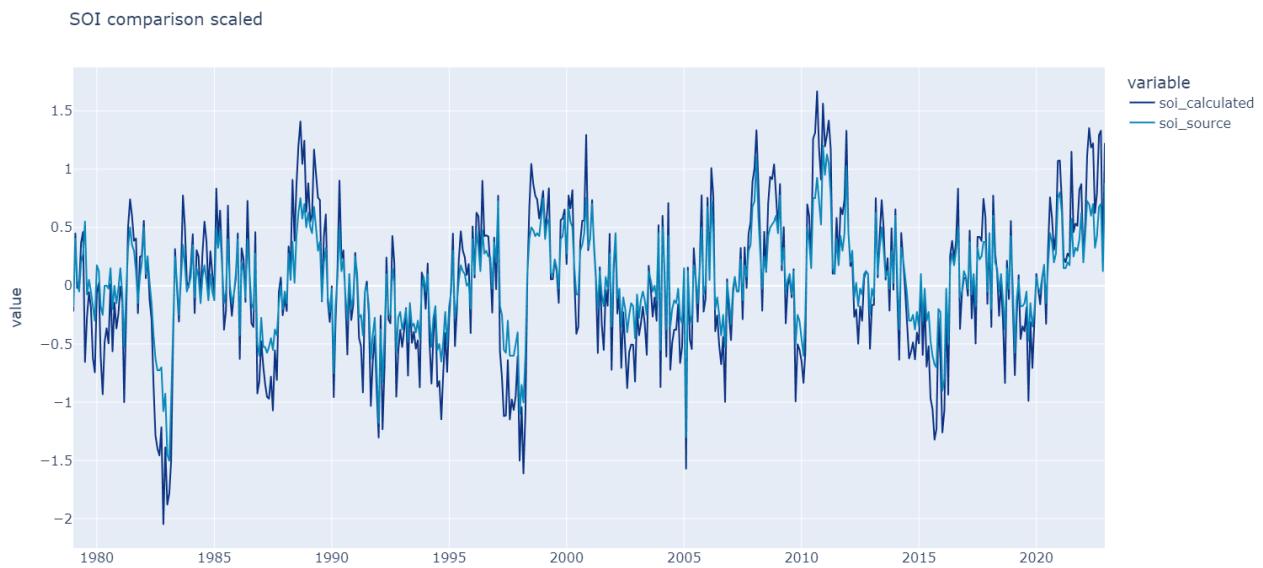


Figure 8 SOI comparison - scaled values

The MJO, AO, and NAO could all be downloaded from their respective sources. No calculations or aggregation was needed because they were provided in a daily interval already. The AO and NAO both contained a missing value, which was filled by means of interpolation. The MJO dataset has missing values during the complete year 1978. Hence the data cannot be filled with mean values or simply be dropped without corrupting the continues time series. The time frame limitation was set to 1979 for all datasets as a result.

To create a custom polar vortex index and detect breakdown events, the below listed variables had to be downloaded from Copernicus API (at the pressure levels of 100 to 10 hPa). Once again, the same time frame was used as for the gathered weather data. Some data condensing was already done during this stage, due to performance reasons and hardware limitations. It was aggregated, firstly into the defined grid resolution of whole longitude and latitude, and secondly to daily values. The resolution was chosen after some experimenting and testing (trial and error). This process results in a geographical resolution of  $2 \times 45 \times 6$  (latitude  $\times$  longitude  $\times$  pressure levels), depicted in Figure 9.

API variable	Feature	Unit	Description
temperature	t	k	The air temperature at the given pressure level
u_component_of_wind	u	m/s	The northward component of the wind speed at the given pressure level
v_component_of_wind	v	m/s	The eastward component of the wind speed at the given pressure level

Table 2 Polar vortex attributes - API references

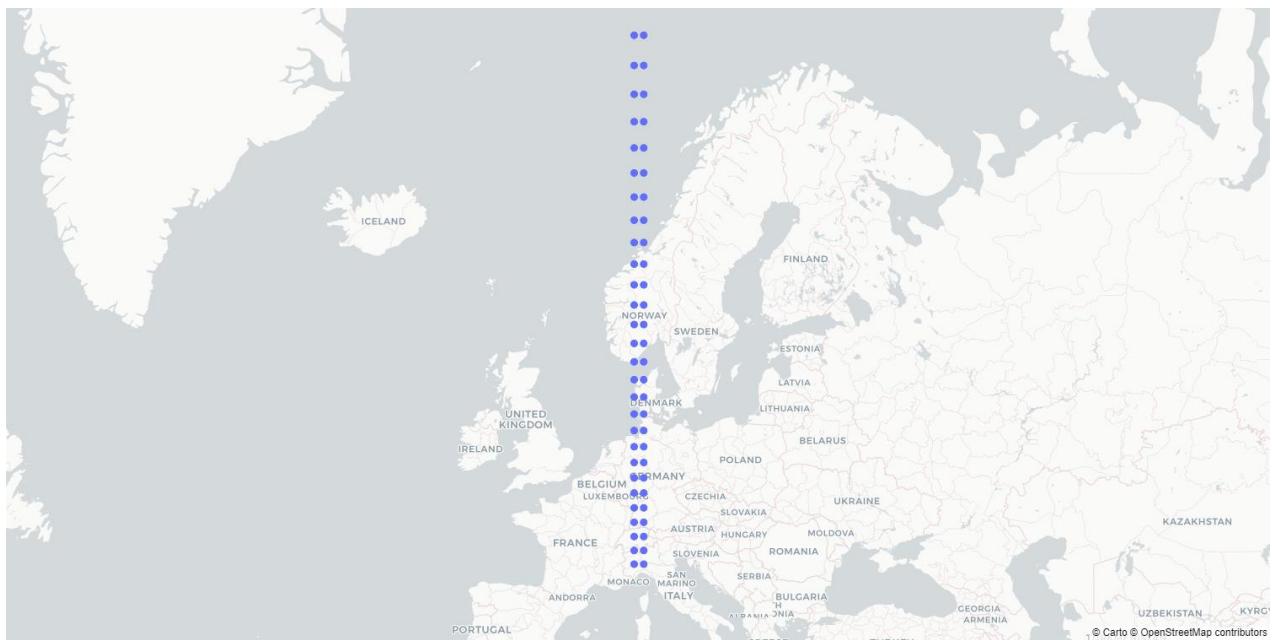


Figure 9 Polar vortex - Data resolution

The five data frames containing the general weather data, the SOI, AO, NAO, and the MJO were joined into one and saved as a csv file. The polar vortex data was saved separately because it was going to be transformed differently. Furthermore, the data was stored in a different format, still containing the longitude and latitude columns.

The merged data frame of the main features (left) consists of the below listed attributes, possessing a shape of 16'000 rows and 15 columns. The secondary data frame, containing the polar vortex (right) data was 9'610'000 rows by 9 columns.

Variable	Unit	Variable	Unit
date	none	date	none
ao	index / none	longitude	degree
nao	Index / none	latitude	degree
mjo_rmm1	index / none	level	hPa
mjo_rmm2	index / none	u	m/s
mjo_phase	index / none	v	m/s
mjo_amplitude	index / none	t	k
soi	index / none	speed	m/s
u10	m/s	direction	degree
v10	m/s		
t2m	k		
cdir	J/m <sup>2</sup>		
sp	Pa		
year	none		
month	none		
day	none		

Table 3 Left to right: Main data frame, polar vortex data frame

## 4.2 Data understanding – Data exploration

Looking at the correlation matrix, nothing too surprising was revealed. Features, such as wind speeds u and v components, and direct irradiation and temperature, possess a high Pearson coefficient. But most of the correlations are negligibly small or even approach zero. As explored in the literature, the NAO and AO do correlate, with a coefficient of 0.5.

Correlation matrix

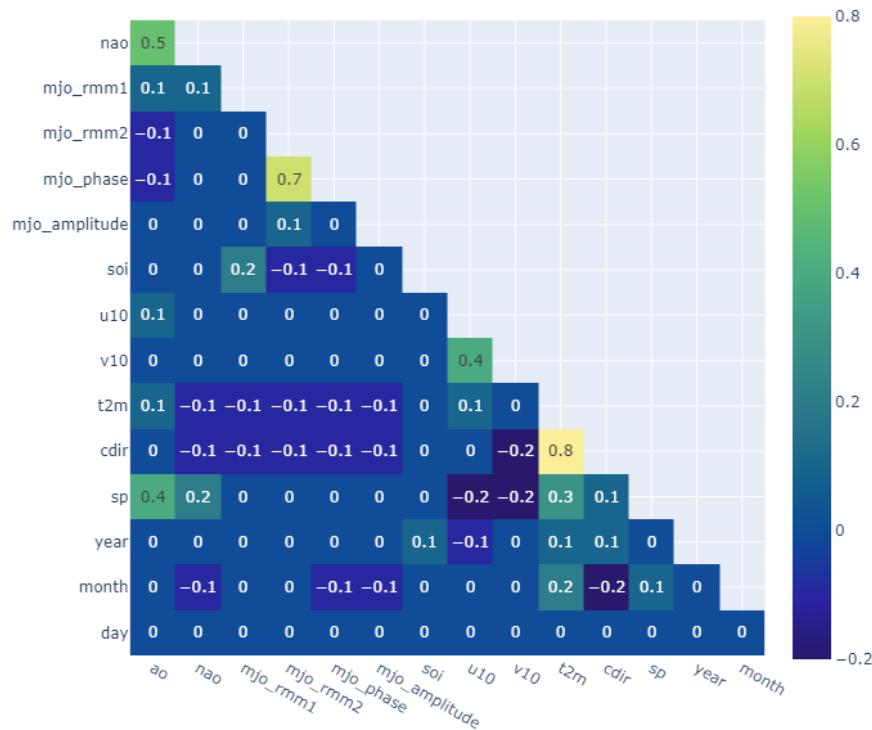


Figure 10 Correlation matrix

Exploring the weather features (temperature, wind speed components, clear sky irradiation, and surface pressure), most of the values behaved as expected. An interesting thing to note in the temperature data was, that the mean is slowly rising, when calculated over the entire time span. It is assumed, that this indicates the effects of global warming. This factor is important to keep in mind because this metric was the target variable of the machine learning models.

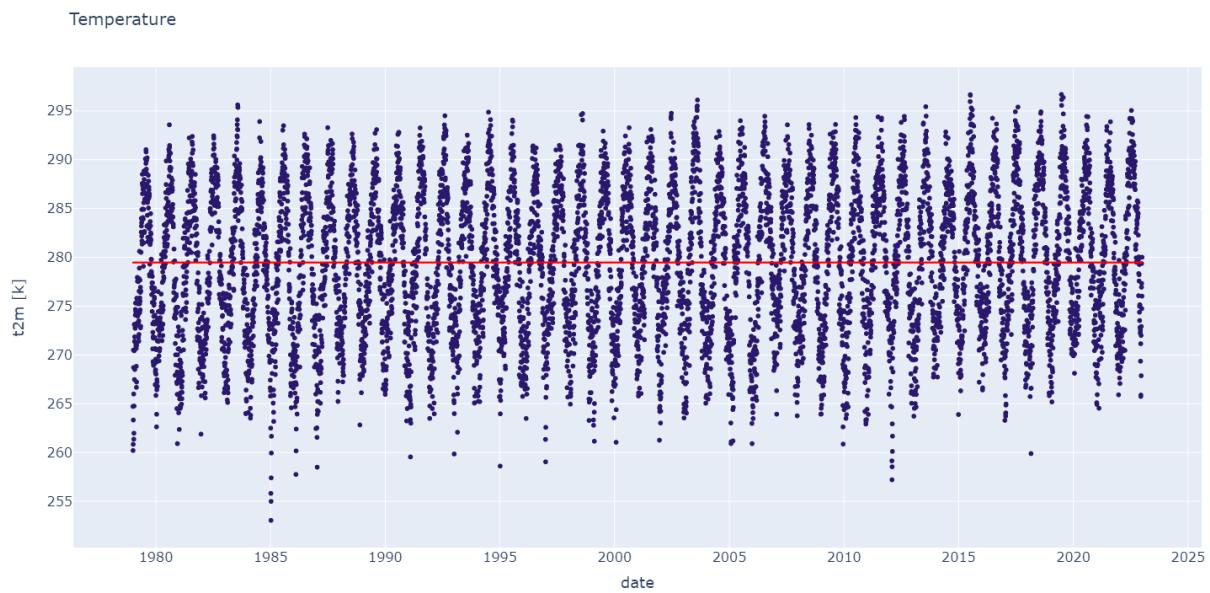


Figure 11 Temperature over time

Further exploring the target variable, the distribution of the values was interesting. The pattern of a multimodal distribution seemed to emerge, with the two peaks being located at the 273 K (ca. 0 °C) and 285 K (ca. 12 °C). Thus, when splitting the data, it is important to keep these factors in mind as well to create an even distribution in the three different sets (see appendix, Figure 26).

As previously seen when validating, the SOI mostly oscillates between the values of + and - 3.5. A hindrance could pose the rapidly shifting value of the SOI. No clear trend can be spotted for long-term forecasting. There is too much fluctuation in the data. To combat this, a rolling mean was layered upon the value (see appendix, Figure 27).

A similar pattern could be spotted with the MJO. Although the values are mostly ranging from 0 to 3, with some outliers. (see appendix, Figure 28).

Once again, the same went for the AO and the NAO. (see appendix, Figure 29, Figure 30).

The custom polar vortex data, consisting of the before mentioned grid containing wind speeds and temperature was quite insightful. A clear change in temperature could be spotted with lower latitudes. Furthermore, most of the time there seems to be two strong winds in the upper atmosphere. It can be assumed, that one is the polar vortex, while the other peak represents the jet streams. But it must be noted, that the data over time, or respectively the wind speeds, are not as consistent as initially expected. The gaps in the plot of the pressure levels came from the data itself. The Copernicus API only provided certain pressure levels (hPa), in this case: 10, 20, 30, 50, 70, 100.

One thing which can be noticed, was the information gained from the grid about major weather situations. Furthermore, when comparing the wind speeds and temperature of a certain time range, the effects of a more unstable polar vortex can be spotted. With lower wind speeds, the temperature around the polar cap increases slightly and spreads further south and thus cooling the air on the lower latitudes (see appendix, Figure 35).

#### 4.3 Data Preparation

After the main data frame was loaded, the cosine and sine functions were applied, by transforming the datetime into a time stamp value. When plotting and overlaying the two functions, it is clearly visible that each data point possessed a uniquely identifiable set of values which oscillate between 0 and 1.

Next the rolling mean was applied to the metrics and features SOI, AO, NAO, MJO amplitude, and temperature (t2m). The time frame of 30 periods / days, relevant for the calculation of the moving

average, was chosen by the time horizon of the forecast. An example of how the moving average applies on the data, can be found in the appendix (see Figure 31).

As already added to the polar vortex data, the wind speed and wind direction were inserted into the main data frame. The distribution of these two values gives some idea on what to expect from the weather itself. The wind mostly comes from north north east and south south east. This concludes that weather and temperatures are carried from the respective directions. The wind speed distribution seems to be a gauss curve, averaging at around 0.75 m/s, which is relatively slow.

The first step of the polar vortex data was to interpolate the missing pressure levels in steps of 10, resulting in 4 interpolations of the values at the 40, 60, 80, and 90 hPa levels. This will aid when analysing the data visually for the detecting of polar vortex breakdown events. Because the interpolated levels lie between available values, the data frame does not need to be split up to prevent falsifying the data.

**Index 1:** When applying the first index and looking at some visual data points, it is apparent that it did not work as expected. It provided some information about the latitude with the lowest irregularities. Although, if the wind speed was considered, it did not yield the desired effect of setting the latitude correctly. It nudged the index into the right direction, but not to the point where it would hold any valuable and consistently correct information. Furthermore, when plotting the index for the temperature spectrum, it did not seem to be of much use as well.

Because of these two factors the second iteration of the index tried to correct those. But trying to combine the two metrics only resulted in an index value, which was sometimes better than the previous version and other times not. Also, shifting the weights into one metric's favour or the other did not significantly improve the index. Reason being, that the temperatures and wind speeds do not line up consistently enough. With these findings, it could be said, that it was not an optimal approach to combine the two metrics into one single index. Hence, it would be more reasonable to provide two separate latitudes or indexes, one representing the wind speed and the other the temperature.

**Index 2:** The second index seemed to suffer from a similar problem. Because the wind speeds are too inconstant over the different pressure levels, the output is not useable. A better result could be achieved with this kind of algorithm, when applied on latitudes and longitudes instead of pressure levels and latitudes. When the algorithm was calculated with the temperature on the other hand, it seems to work quite well for identifying a suitable and shifting latitude, which represents the temperature front in some way.

The additionally tested idea of a threshold which has to be exceeded and the rolling mean detection of a breakdown event can be applied on other indexes as well. The detected polar vortex breakdown events with this method seemed applicable, depending on how much time was put into optimizing for the optimal windowing period, sensitivity and threshold. The problem was, that in order to optimize those values, all 16'000 data point would have to labelled, whether they represent a breakdown event or not. This was not feasible for two reasons: The given time frame and the unavailability of a conclusive data set or table listing all smaller breakdown events. If there is such a data source, the author was not able to find it.

**Index 3:** Because the third index did not work on a latitude level, the before mentioned idea of detecting breakdown events cannot be applied to this algorithm. Nonetheless, this way of obtaining information of the polar vortex showed one major problem when used for wind speeds. There are multiple different occurrences of higher winds speeds in the data, which do not necessarily indicate a polar vortex. Such higher values do falsify the validity of the calculated delta. But when applying the same index on the temperature metric, the result hold more information. But no clear pattern could be detected in the index, when plotting it over a time axis.

**Index 4:** Although it is the simplest index, it seemed to work the best when it came to identifying maximum wind speeds, considering the given sample. It also circumvented the problem of multiple different peaks on different pressure levels. But most certainly, this problem will be prevalent with all indexes, especially if there is no clear maximum which can be identified. But when using the index on the temperature metric, most often it only pointed out the most north latitude. Therefore, not providing much useful information.

During further optimization of the indexes 2 (temperature) and 4 (wind speed) for their respective metric, the input values were changed around, and the results were checked visually by checking the time frames of the predicted breakdown events. Additionally, for both used indexes, only breakdown events were considered, which happened during the months December until April. An example of such a plot, representing a breakdown or weak event can be found below. It clearly depicts the behaviour described in the literature (see chapter 2.5). In the beginning the wind speeds are higher and the cold air is trapped and clearly separated from the warmer air. With the following decrease in wind speeds, the clear-cut border of the temperature vanishes, and the temperature gets evened out among the latitudes. This results in cooler air temperatures in the southern regions.

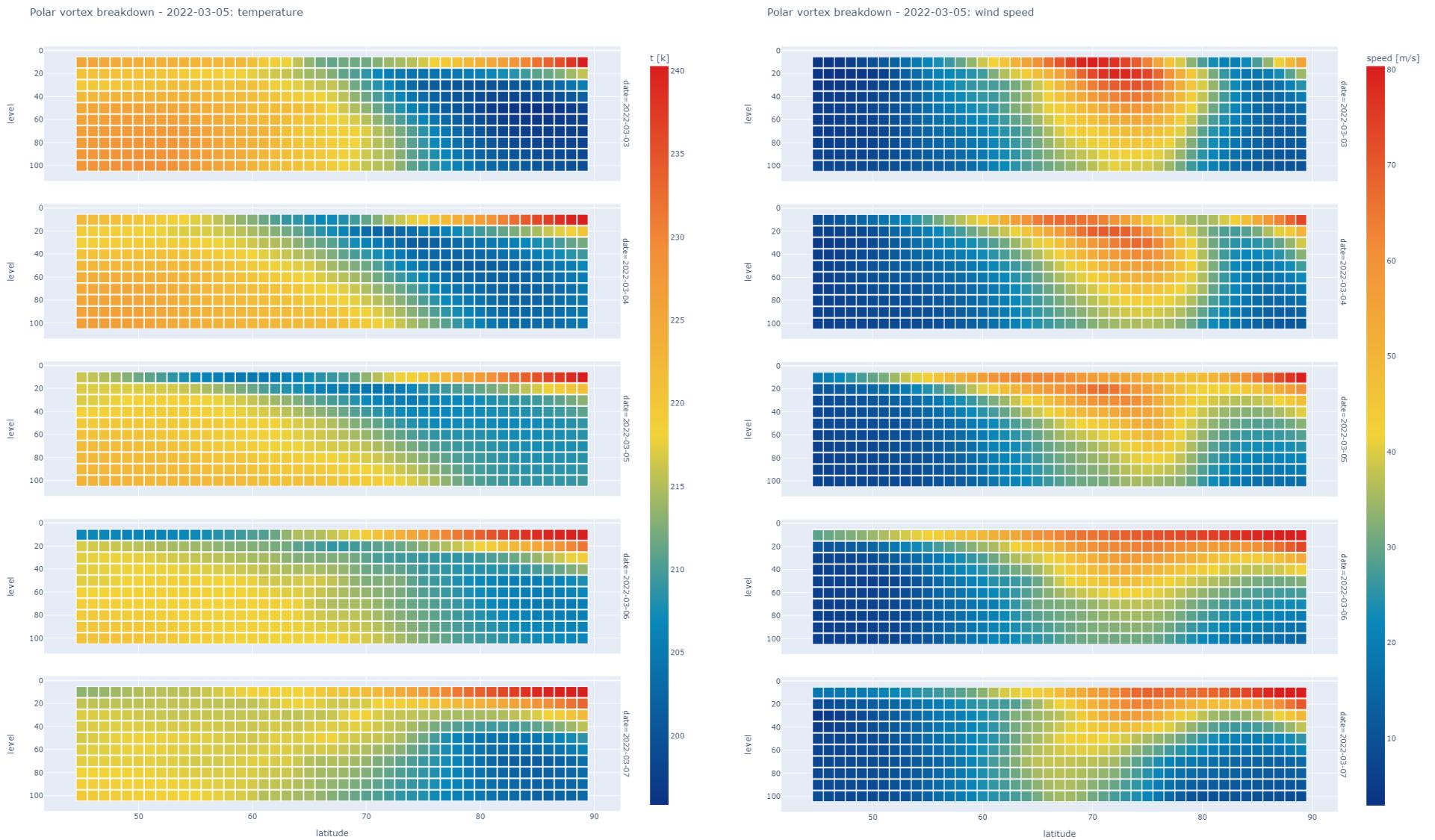


Figure 12 Polar vortex breakdown event: left index 2, right index 4

With the given results on each index and some additional testing, the following indexes were concluded to work the best for the respective metric, in conjunction with the given parameters:

Metric (code variable)	Index	Parameters	
Temperature (PVT)	Index 2	n_lat	= 3
		threshold	= 1.5
		break_down_offset	= 15
		break_down_sensitivity	= 0.1
Wind speed (PVS)	Index 4	threshold	= 0.5
		break_down_offset	= 60
		break_down_sensitivity	= 0.875

Table 4 Index and applies parameters

Each of these indexes added the following features to the main data frame:

- latitude: The latitude of the polar vortex edge, according to the index
- mean: The mean of the respective metric at the given latitude
- pv\_edge: True, if there is an edge (threshold met); False, if there is no edge (threshold not met)
- pv\_break\_down\_event: True, if a breakdown event is detected; False, if there is no event

The two prefixes pvt (polar vortex temperature, index 2) and pvs (polar vortex speed, index 4) were provided in the data frame to distinguish the features. Comparing these values, some correlations can be seen between the two, especially on the latitude level. Although, it could be improved, shown by the outliers setting the latitude at the end of the available spectrum. Apart from that, it can be induced, that the two polar vortex indexes, set the border at around the same latitude.

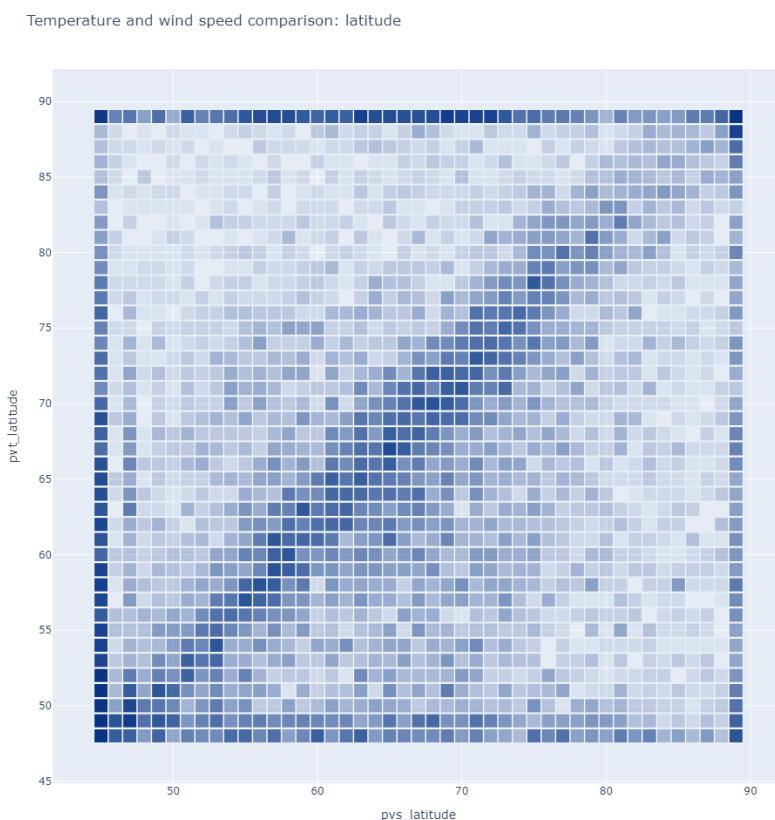


Figure 13 PVT and PVS latitude comparison

Considering, that the breakdown event parameters are somewhat different, it is not surprising to see that the breakdown events themselves do not correlate with one another as much. Furthermore, both seem to have a higher number of events during the month April, which makes sense to some degree, taking into account the explored literature. Additionally, only the temperature-based index set breakdown events to occur during December. A reason for this could be the greater lag of the wind speed-based index breakdown events. The overall number of events is similar with 23 for the temperature, and 21 for the wind speed-based index (see appendix, Figure 32).

To further verify the indexes, each of them was compared to the SSW events. Some of the months match up with the breakdown events, but the frequency is too low to be conclusive. Thus, further work on the custom indexes would be needed, or at least the breakdown detection. As previously stated, a needed addition in verifying the indexes would be a conclusive list of the weak polar vortex events.

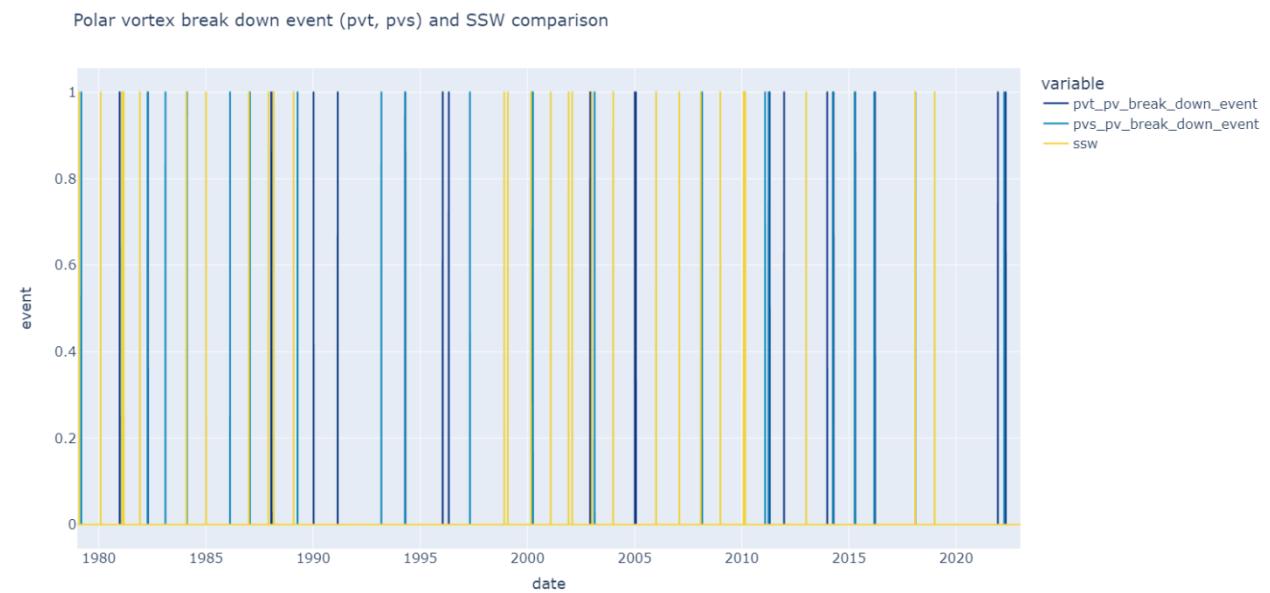


Figure 14 Polar vortex breakdown and SSW comparison

But more important than to match the polar vortex breakdown events to the SSW events was the comparison and correlation to the outdoor temperature ( $t_{2m}$ ). Both of the latitude values have a Pearson coefficient of -0.3, concluding that some linear correlation between the mean temperature and the calculated indexes exists. Furthermore, the data distribution, when separating by edge detection, reveals a different density in certain areas.

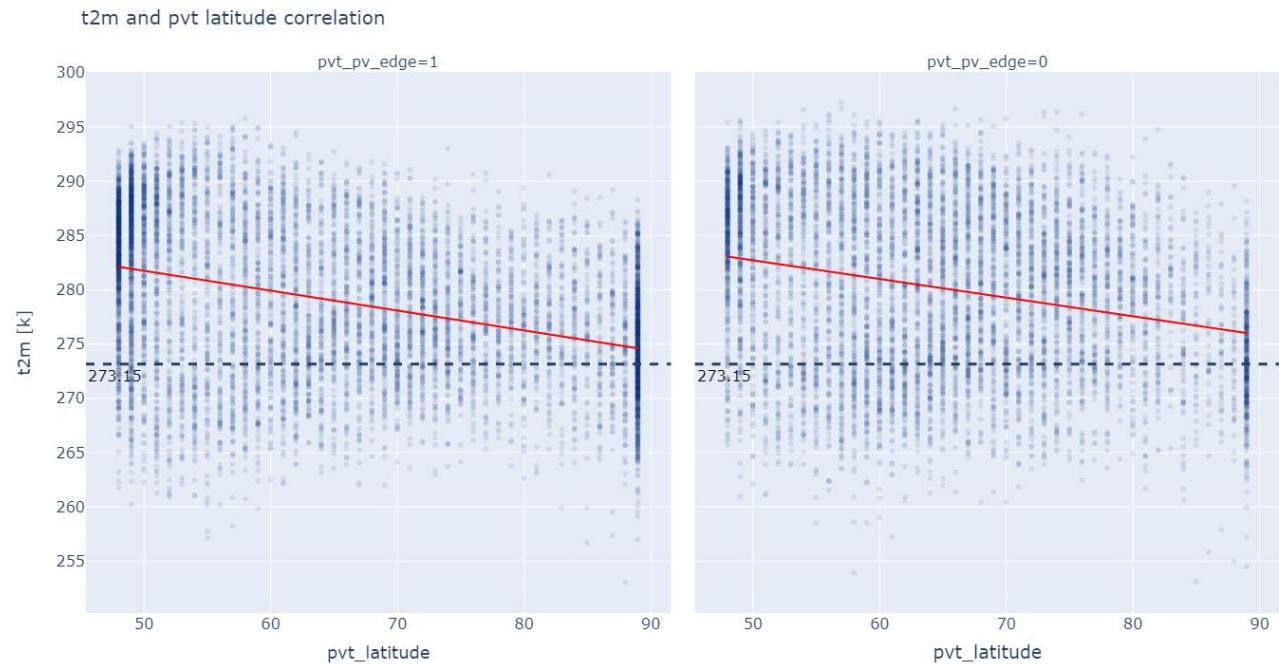


Figure 15 Index latitude and temperature corelation

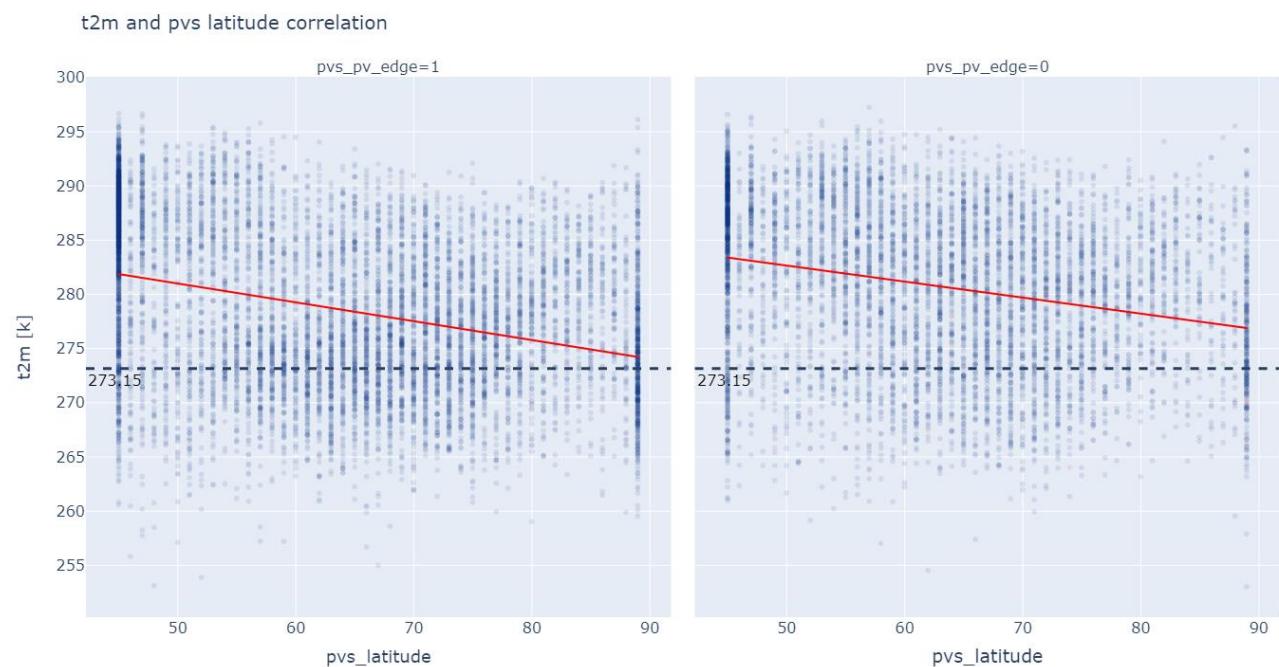


Figure 16 Index latitude and temperature corelation

Lastly, the target vectors were created. The two periods t1 and t2 got calculated and added to the dataset. Afterwards, the mean value of the t2m, grouped by day of the year was calculated across the whole data set to generate the classes. A classification based on a monthly median (calendar month) is not a valid approach. During spring and fall, the beginning and end are vastly different temperature wise. Thus, such a classification would lead to a heavy bias and falsify the results. Instead, a mean for each day was calculated, against which the current value was compared.

Because the classification is done based on the daily mean and not median, there is not an exact 50/50 split in the classes. But the gained advantage is a much smoother average temperature across the year, with which the datapoints were classified. If a value was above the mean, it was classified as 1 and 0 if not.

The outlier on the temperature curve can be explained by the amount of data which is available for the respective datapoint. In this case meaning, that the outlier falls on the 29<sup>th</sup> of February. Thus, having 4 times less data points available. An interpolation would be possible, but since this is not a mistake, it will be kept as is. Either way, the impact of this change would be minuscule.

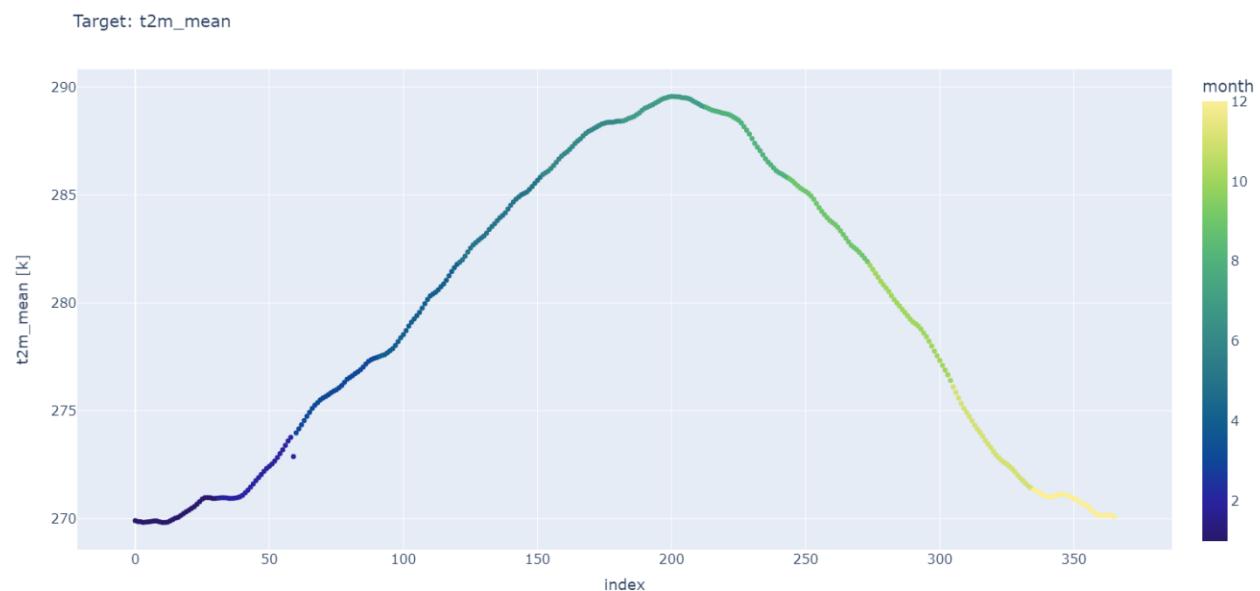


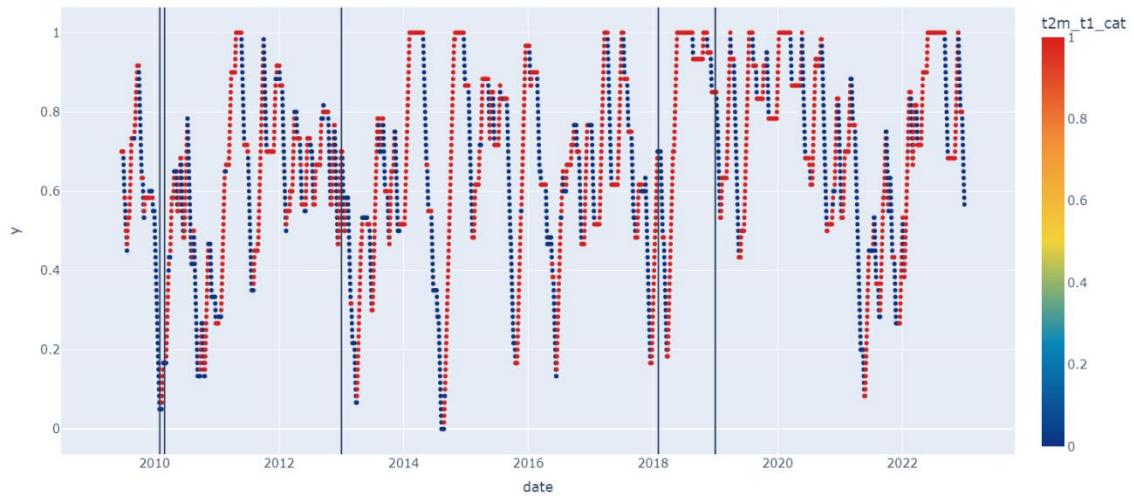
Figure 17 t2m mean values

After creating the target vector, the two categories 1 and 0 can be compared to the three breakdown events of SSW, PVT, and PVS, to check if they yield any value for a classification model. The plot shows the category over the time axis, with an applied rolling mean of 60 to better visualize shift in trends. Moreover, not all datapoints were illustrated, due to spacing issues. The breakdown events are signalled with vertical lines.

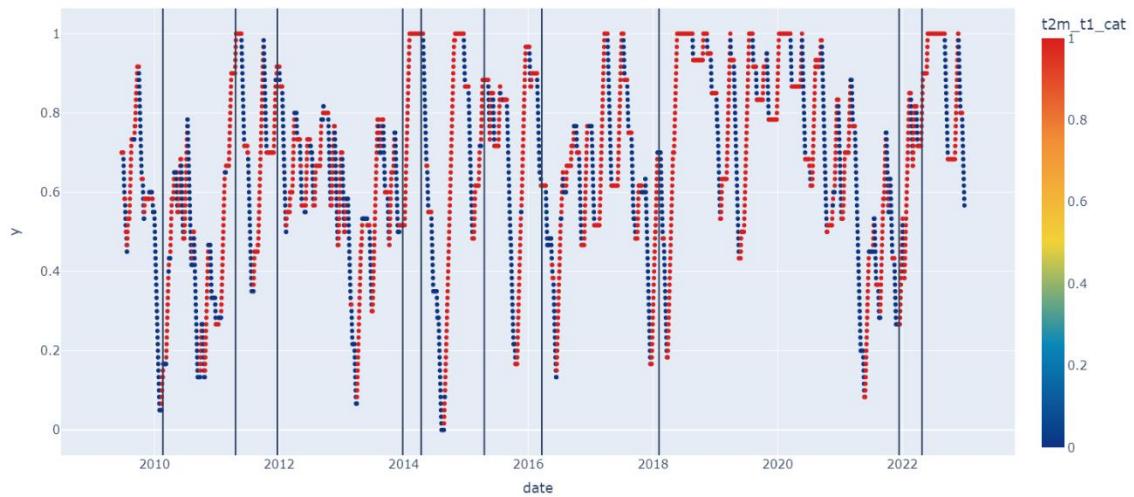
The breakdown events of the PVS and PVT were placed similarly in the time series, although the later one detected more events during the inspected period. Nonetheless, it seems that in both cases, more often than not a period of below average temperature days follows a predicted or detected breakdown event.

Furthermore, the SSW events follow a similar pattern with the below average days, although it is more infrequent during the compared years and does not line up as much with the other two breakdowns. It can be concluded that the breakdown detection works to some degree in predicting a period of below average temperature days. But, as previously stated, the method could be further optimized.

Break down event: ssw



Break down event: pvt\_pv\_break\_down\_event



Break down event: pvs\_pv\_break\_down\_event

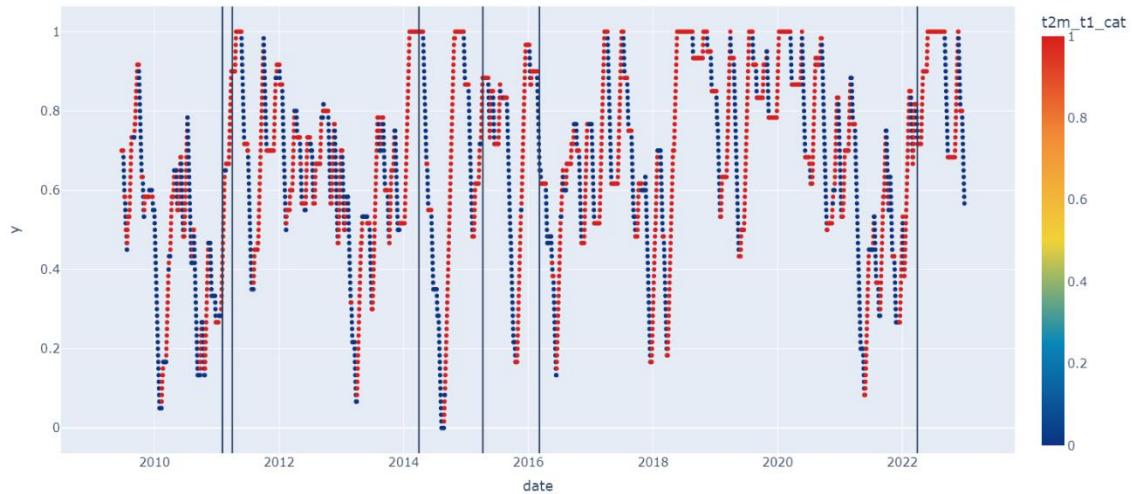


Figure 18 Breakdown event and category comparison

The categorical values seemed to occur in bigger trends, meaning that a category is much more likely to be followed by the same category immediately after, rather than the opposite. Thus, it could be assumed, that the target vector would hold much more values, where the t1 category is equal to t2. But it has to be reiterated, that the t1 and t2 categorical target vector were spaced by 14 days. When comparing these two cases ( $t1 == t2$ ;  $t1 != t2$ ) it can be stated that it is more likely for the categories of the target vector to be the same, but not by a huge margin (same category  $p \approx 0.55$ , not same category  $p \approx 0.45$ , see appendix, Figure 33). This fact is relevant when evaluating the model performance. If an equal distribution of values across the 2-dimensional target vector is assumed ( $q = 0.5$ ), the overall random guess accuracy would come out to be  $1/2^2$  or 0.25. Hence, if the model does not converge and randomly guesses a category, the accuracy will come out to be 0.25.

#### 4.4 Modelling – Classification Models

##### 4.4.1 Data Preparation

To prevent any data loss, the application of the shifting window was done before splitting the data frame into the training, validation, and testing set. The rolling past 30 days were added as new features (the shifting window), with a suitable prefix to identify the data. This was only done on the input features. The past target vector variables were dropped from the x (input) data frame, to prevent any future information to get into the features and to prevent the problems stated in the previous paragraph. Plus, for the classification models the t2m\_t1 and t2m\_t2 were dropped and for the regression models the t2m\_t1\_cat and t2m\_t2\_cat.

When splitting the data, two things needed to be considered: The temporal integrity and the distribution of the target categories. Hence, the validation and testing set cannot be taken from the end of the data frame. Reason being the clear shift to above average data points. Instead, the training set was kept temporally intact, and the validation and testing set were taken from the beginning and end of the data frame, leading to an overall even distribution (see appendix, Figure 34). Finally, the standardization was applied to all datasets, based on the mean and standard deviation of the training set.

##### 4.4.2 Random Forest

For the random forest model, the below listed parameters were optimized, with the given ranges. The nonlinear scaling of most optimized features was chosen to reduce processing / training times, while retaining a wide range of tested values. All 180 permutations of the below listed values were tested. All model parameters can be referenced in the accompanying notebook.

Parameter	Scaling	Range
n_estimators	$2^n$	2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096
max_depth	10n	10, 20, 30
min_samples_leaf	$2^n$	2, 4, 6, 8, 10

Table 5 Random forest parameter tuning

After optimizing, the following parameters performed the best on their respective metric. Noteworthy is, that the model with the best overall accuracy, also performed the best on the t1 metric, which is not the case with the t2 category. One problem, which both models share, is, that they overfitted on the validation set. Although this is the case and is counteracted to some degree by the high parameter value of min\_samples\_leaf and max\_depth, it does exceed the set goal of 0.25 and 0.5 accuracy on all sets. The optimal model parameters are given below. Moreover, the model does not seem to have a strong categorical bias. This would be apparent, when looking at the confusion matrix. The labels among the FN and FP are more or less evenly distributed, although there seems to be slight empathies on the FP.

Optimal parameters:

- n\_estimators: 128
- max\_depth: 10
- min\_samples\_leaf: 8

Set	Accuracy t1	Accuracy t2	Accuracy
Train	0.956	0.956	0.915
Valid	0.703	0.665	0.501
Test	0.649	0.629	0.429

Table 6 Optimized random forest accuracies

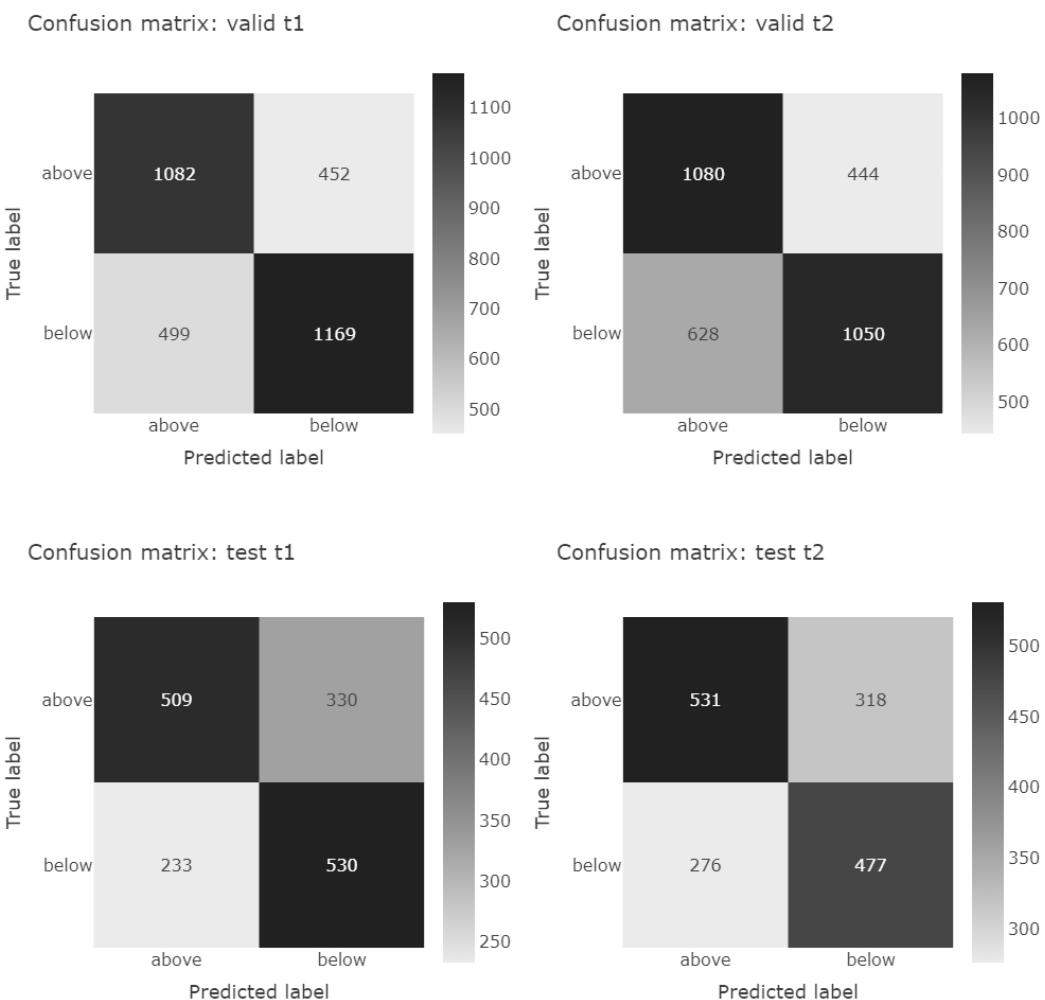


Figure 19 Random Forest confusion matrix

Because the model has some skill, more information was gained by evaluating the feature importance over each feature and over the shifting window. It seems, that the model has recognized the importance of the weather phenomenon indexes and their impact on the average temperature. As depicted, the rolling means of the AO, NAO, MJO, SOI are among the most important features. Because of the shift in the categories over the used time span, the year is the most important feature.

Unfortunately, the pvs and pvt breakdown event indicators were not adapted as a feature by the model. This makes sense to some degree. The given information could be derived by the model itself, as all pvs and pvt related values are based on the pvs\_speed\_mean and pvt\_t\_mean respectively. Nonetheless,

one value of the custom indexes is ranked in the top half, when it comes to their importance: The pvs\_speed\_mean, which tracks the average speed at the centre of the assumed polar vortex. The pvt\_t\_mean, which represents the temperature in the same way as the pvs\_speed\_mean is listed not too far below. This concludes that the engineered indexes hold some relevant information, but could be improved, as already suspected.

Comparing the importance of each time offset, an almost even distribution can be seen, when neglecting the first three offsets. These are weighted higher than the rest, although not by a significant margin. Because there is no drop off on the bigger offsets, even more past data could be considered important for the prediction.

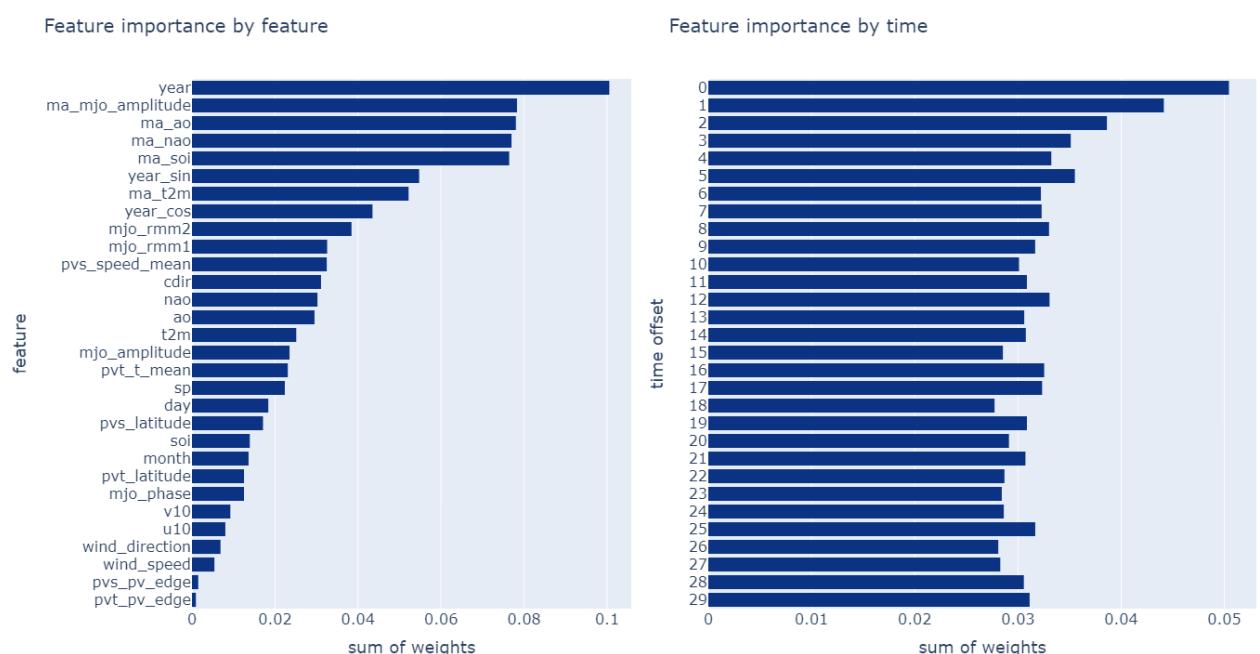


Figure 20 Feature importance by feature and time

#### 4.4.3 Nu-support Vector Machine

The Nu-support vector machine was provided with too many datapoints and dimensionalities. This concluded in extensively long calculation times for even a single model. Trying to implement multiprocessing did not improve the performance significantly, making it not possible to optimize this model and therefore create meaningful insights.

#### 4.4.4 Multi-layer Perceptron

The main factor, which was optimized and has a huge impact on the model, is the architecture, or the hidden\_layer\_size parameter. The other parameters were fixed and not changed to avoid a large number of permutations which would have to be tested. The alpha value is fixed to 0.1 for the architecture optimization. Once again, the number of neurons were not scaled linearly to minimise the number of permutations to run and test. The below given range was tested, resulting in 780 different architectures. After finding the optimal architecture, the listed alpha values were tested as well.

Parameter	Scaling	Range
n_layers	n	1, 2, 3, 4
n_neurons per layer	n_features * / 3 <sup>n</sup>	0, 12, 36, 110, 330, 992
alpha	nonlinear	1, 0.5, 0.25, 0.1, 0.01, 0.001

Table 7 Multi-layer perceptron parameter tuning

The optimal hidden layer size of the network is interesting. The input layer consists of 992 nodes, but the following layer only possess 12, meaning the data is heavily condensed down. On the last hidden layer, the inputs are extrapolated again before a class is predicted. The optimal architectures for t1 and t2 show a similar shape. It can be assumed, that with the given data such an architecture is beneficial. Architectures with higher number of neurons tend to overfit much more onto the training data. Theoretically, this could be counteracted by increasing the alpha value for networks with higher number of neurons. A scaled down depiction (by factor 4) of the optimal architecture can be found in the appendix (see chapter 11.4). Similar to the random forest model, the predictions seemed to favour the below category, as there are significantly more FN predictions than there are FP. Counterintuitively, the t2 accuracy on the test set is higher than the t1's.

Optimal parameters:

- hidden\_layer\_size: [12, 12, 12, 330]
- activation: relu
- solver: adam
- alpha: 0.1
- learning\_rate: adaptive
- shuffle: False
- early\_stopping: True
- max\_iter: 200

<b>Set</b>	<b>Accuracy t1</b>	<b>Accuracy t2</b>	<b>Accuracy</b>
Train	0.883	0.823	0.734
Valid	0.652	0.651	0.448
Test	0.559	0.624	0.366

Table 8 Multi-layer perceptron accuracies

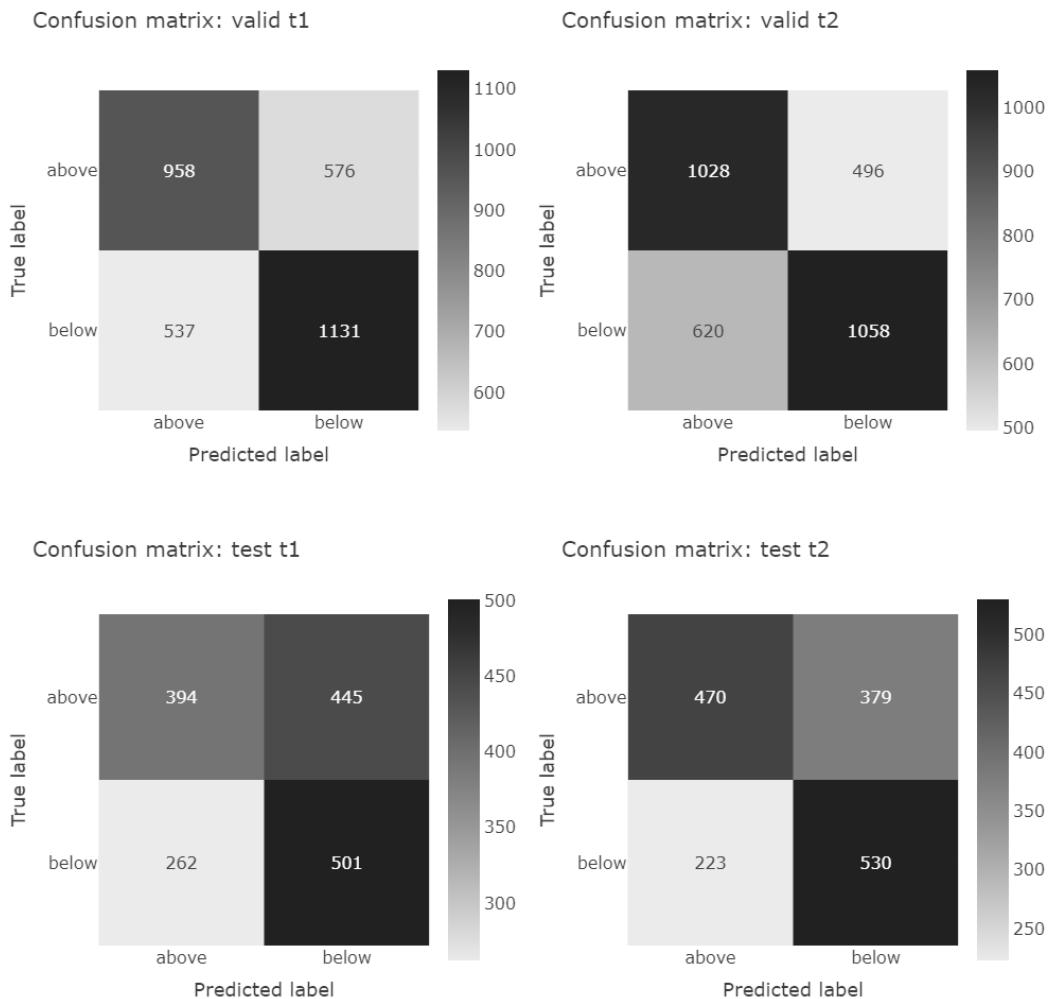


Table 9 Multi-layer perceptron confusion matrix

#### 4.4.5 Recurrent Neural Network

The main problem which arose during the training of the models, was the time and resources each model used up. Because of this, only a handful of models were trained, and the optimization process was cut short. The RNN consisted of a dynamically scaling lambda input layer, bidirectional hidden layers with LSTM-nodes, and a dense output layer. Early stopping and a low number of training epochs (5) were implemented to reduce training times. The optimized parameter of the model was limited to the following architectures:

Parameter	Scaling	Range
hidden layers	nonlinear	[496], [248], [165], [496, 496, 496]

Table 10 Recurrent neural network parameters

Because of the before mentioned problems during training, in combination with the worse accuracy in comparison to the RF and MLP models, the running of the optimization has been quit after these four models. Although, the reviewed literature states, that the RNNs are very suitable for time series prediction and were widely applied, it does not seem to have the desired output. There are three possible reasons for these circumstances:

- The hidden layer sizes or other parameters were not chosen correctly and would have to be tested more thoroughly
- The data is not suitable for this type of model
- The classification is not the right way to utilize a RNN with time series data

In order to prove any of these statements, more research on this specific topic in combination with the prepared data and problem would have to be conducted.

Optimal parameters:

- hidden\_layer\_size: [496, 496, 496]
- activation: selu
- solver: SGD (Gradient descent)
- learning\_rate: 0.0001
- early\_stopping: if binary accuracy < 0.001
- epochs: 5
- loss: Binary cross entropy

Set	Accuracy t1	Accuracy t2	Accuracy
Train	0.512	0.489	0.220
Valid	0.521	0.476	0.176
Test	-	-	-

Table 11 Recurrent neural network results

#### 4.5 Modelling - Top Model Optimization (RF)

After evaluating the different model types, the top model (the RF classifier) was further explored by the means of dimensionality reduction and preventing the overfit. As previously seen in the feature importance plot of the model (Figure 20), there are some features which barely hold any weight in the predicting. Thus, 4 new data sets were created, each with a different set and number of features. Set 3 was created to see how the model would perform with only the moving average values and neglecting the raw data with more volatility.

- Set 0: Top 5 features according to the feature importance
- Set 1: Top 10 features according to the feature importance
- Set 2: Top 15 features according to the feature importance
- Set 3: All features except nao, ao, t2m, mjo\_amplitude, sp, day, soi

Set 2 performed the best with a validation accuracy of 0.474, being slightly worse than the full data set. This seems like an acceptable trade-off between model performance and complexity. The problem with overfitting could not be solved by the feature reduction.

Another optimisation run was created, with the complete data set and the below listed model parameters. The main goal was to create a model which performed approximately equal on the training and validation set, while retaining its previously established predictive skill.

Parameter	Scaling	Range
n_estimators	fixed	128
max_depth	2n	1, 2, 3, 4, 5, 6, 7
min_samples_leaf	25n	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Table 12 Random forest classifier overfit parameters

The top model was then picked by plotting the validation accuracy against the training accuracy. The optimal model sits where those metrics are approximately equal and before a drop off in the validation accuracy, which indicates the overfitting (see Figure 37). This resulted in the following parameters and model performance. The accuracy on the test and validation was improved and the overfitting does no longer occur. Meaning, the model generalises better than its previous iteration.

Optimal parameters:

- n\_estimators: 128
- max\_depth: 8
- min\_samples\_leaf: 225

Set	Accuracy t1	Accuracy t2	Accuracy
Train	0.726	0.693	0.501
Valid	0.727	0.73	0.552
Test	0.665	0.66	0.473

Table 13 Random forest classifier performance (run 2)

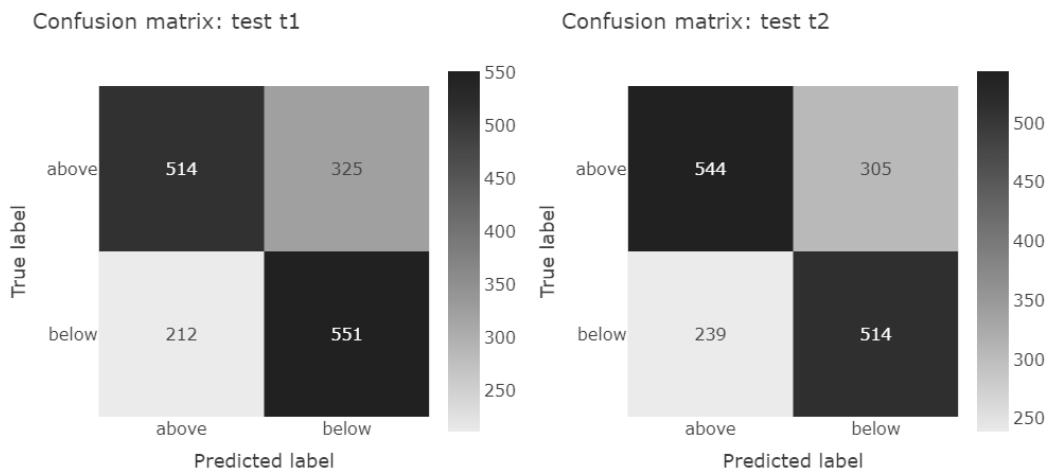


Figure 21 Random Forest confusion matrix

#### 4.6 Modelling – Regression Model (MLP)

To propose a similar model and comparable approach to the reviewed literature, one regression model was created. The chosen model was a MLP regressor. Because the study used as a benchmark [32] also made use of artificial neural networks, this should yield a comparable approach and results. A comparison on the top classification model, the RFC, would also have been valid.

The data was split and standardized the same way as before with the classification models. Additionally, the categorical columns were dropped (t2m\_t1\_cat, t2m\_t2\_cat). Because the classification was the main

goal of this thesis, only a handful of architectures were tested on the regressor. All the other model parameters were fixed to the same values as the MLP classifier.

Parameter	Scaling	Range
hidden layers	nonlinear	[496], [248], [165], [496, 496], [248, 248], [165, 165], [496, 248], [248, 496], [496, 496, 496], [248, 248, 248], [165, 165, 165], [496, 248, 124], [124, 248, 496], [496, 496, 496], [248, 248, 248, 248], [165, 165, 165, 165], [496, 248, 124, 62], [62, 124, 248, 496], [496, 496, 496, 496], [248, 248, 248, 248, 248], [165, 165, 165, 165, 165], [496, 248, 124, 62, 31], [31, 62, 124, 248, 496],

Table 14 Multi-layer perceptron parameters

This resulted in the below mentioned best performing model. To make the outputs and thus the RMSE comparable to the literature, the predicted vector, alongside the true values, were unstandardized by reversing the given formula. The optimal parameters can be found below. A scaled version (factor 25) of the network architecture can be found in the appendix (see chapter 11.4). When comparing the unstandardized results with the benchmark of 2.117 °C, the provided model performs slightly worse, depending on which set is considered. This means, that there would be some room for improvements, if the focus would lie on the regression models. Two noteworthy findings are, that firstly an architecture with 4 layers was preferred, rather than 3 like on the MLP classifier. And secondly the shape of the architecture is cone like, and slimmer before the output layer. Hence, a completely different type of architecture is needed to perform well on a similar problem.

Optimal parameters:

- hidden\_layer\_size: [496, 248, 124, 62]
- activation: relu
- solver: adam
- alpha: 0.1
- learning\_rate: adaptive
- shuffle: False
- early\_stopping: True
- max\_iter: 200

Set	RMSE t1	RMSE t2	RMSE	R <sup>2</sup> t1	R <sup>2</sup> t2	R <sup>2</sup>
Train	0.567	0.644	0.607	0.994	0.992	0.993
Valid	2.184	2.374	2.267	0.905	0.891	0.898
Test	2.670	2.508	2.590	0.863	0.878	0.871

Table 15 Multi-layer perceptron results (unstandardized)

#### 4.7 Modelling – Polar Vortex Clustering

In order to process the data as grey scale images, the tabular values were restructured into 2d arrays, representing the daily situations. The x-axis represents the latitude and the y-axis the pressure levels. All pixels were standardized to make the values more comparable. The data was split randomly, with a fraction of 0.8 as training and 0.2 as testing data set. Because this was an unsupervised approach no validation set is needed. Due to the fact, that this should only show the possible applications for image clustering in weather forecasting with polar vortex data, only the wind speed metric was considered.

The KMC (implementation from the sklearn library) was fit to the data with the below listed parameters. Those were chosen randomly and based on the given library documentation. Some alterations were made during the creation process of the model to optimize the values. But further work would be needed and other methods to be tested.

- n\_clusters: 9
- algorithm: full

A sample of pictures from each cluster can be found in the appendix (see chapter 11.3). Looking at the distribution among the clusters (see appendix, Figure 36), a very uneven picture can be spotted, with the majority (0.45) of the images being labelled as cluster 1. This distribution was retained when changing the number of clusters. This cluster contained images with almost no windspeed at all.

After a visual analysis of each cluster, the following statements can be made about each:

Cluster	Situation
0	Weak polar vortex, clear outline
1	No polar vortex, almost no wind speeds
2	Strong polar vortex, well defined outlines
3	Moderate polar vortex, loosely defined borders
4	No polar vortex, moderate wind speeds
5	No polar vortex, slow wind speeds
6	Weak polar vortex, unclear outline
7	Moderate polar vortex, clear outline, lower latitudes
8	Moderate polar vortex (split), two outlines

Table 16 Cluster description

When grouping each cluster by month, the cluster 2 occurrences were much more frequent during the summer months. Which can be expected, as there are almost no wind speeds in the observations for said cluster. Each of the other cluster shows a different distribution among the months. Henceforth, it can be assumed that the number of clusters is fitting to some degree for the assessment of temperature anomaly detection (see appendix Figure 38).

#### 4.8 Gas Price and Outdoor Temperature Correlations

The gas prices, represented by the spot price in USD for natural gas per million Btu was downloaded from the given source, in a daily resolution [77]. Said prices as well as the t2m temperature were both standardized to make them comparable. Additionally, the previously calculated categories of the t2m value were also regarded as relevant and added to data frame.

Because the gas price underlies many other influence factors it was hard to tell, how directly the temperature affects it, by only looking at the price evolution over time. To create a detailed analysis, a list of major impacting events would have to be compiled and regarded when drawing conclusions.

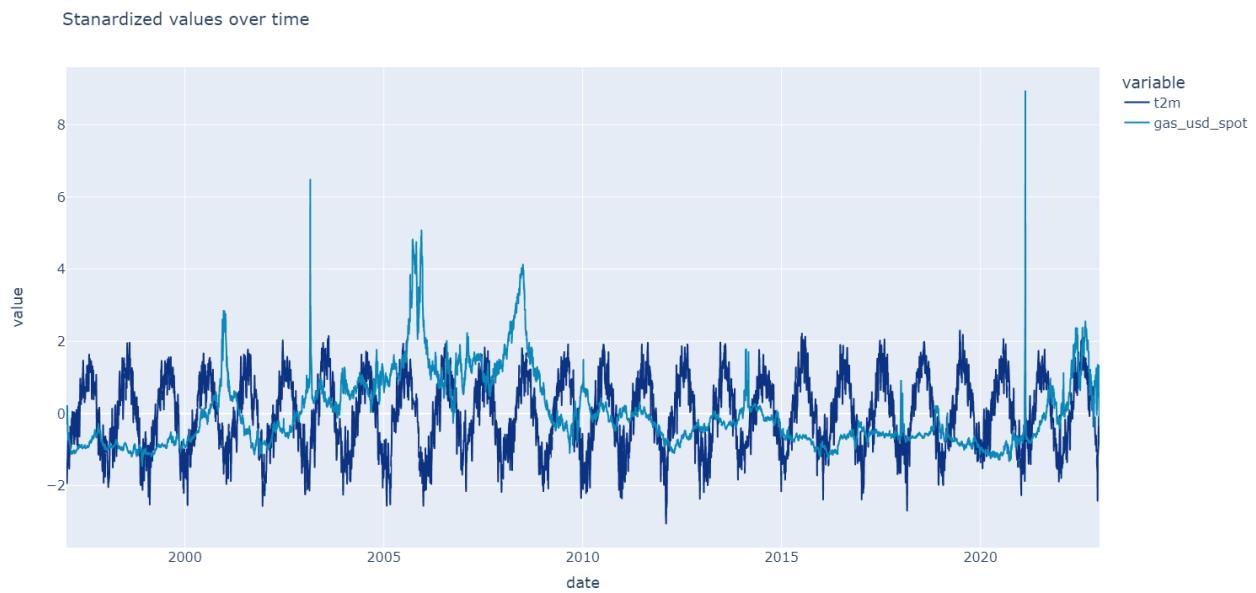


Figure 22 Standardized gas price and t2m

A first look at the correlations, split up by the two categories of above and below average days did not yield any results. The correlation coefficient approaches zero. No clear visual correlation seems to emerge when plotting the data. Only analysing the colder months (October to March), during which households consume heating energy, did not convey more data. The correlation is slightly higher, but not to a significant degree. As expected at this point, the change in gas price as a decimal value also shows no correlation to above or below average days, even if only the heating months were regarded.

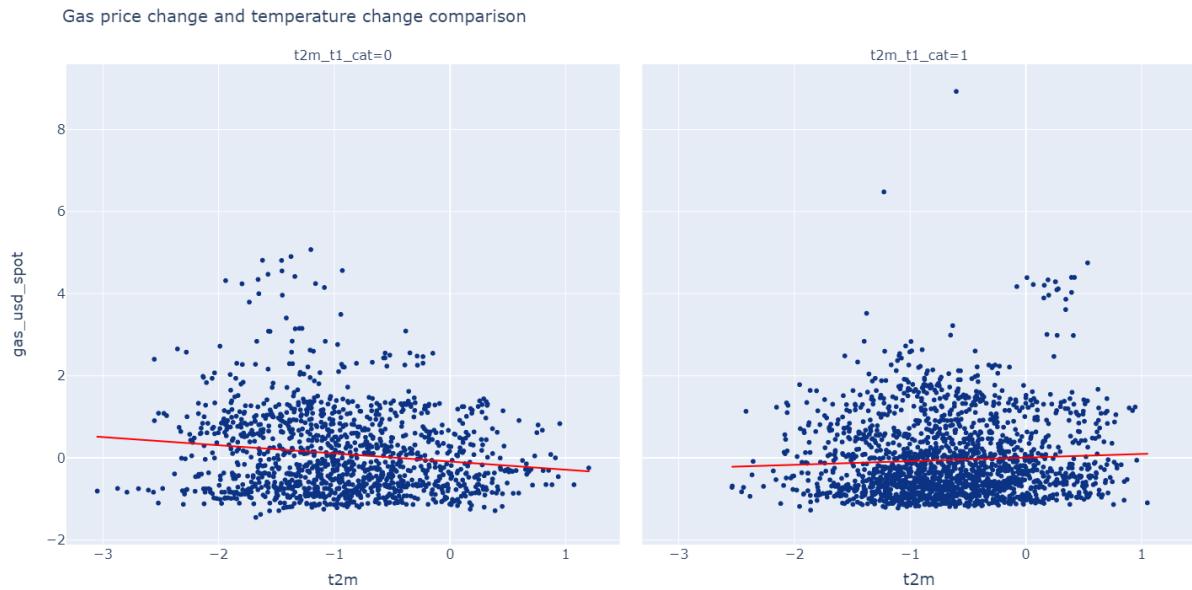


Figure 23 Gas price and temperature correlations (scatter)

Overlying the standardized temperature and natural gas price did not produce any insights. Comparing the two time series data points is not easy, because of their different nature. The gas prices are very irregular with high peaks at seemingly random intervals, whereas the temperature is highly cyclic in nature. In order to negate this factor, two peak identifications methods were applied to the temperature data:

- Robust peak detection (t2m\_zscore)
- Manual threshold value (t2m\_th)

The identified peaks were plotted over the gas price development. On one hand it could be argued that some of the identified peaks correlate with a rise in gas prices. But on the other hand, this only seems like a loose connection. There are also peak events, with absolutely no upward driving effect on the gas prices.

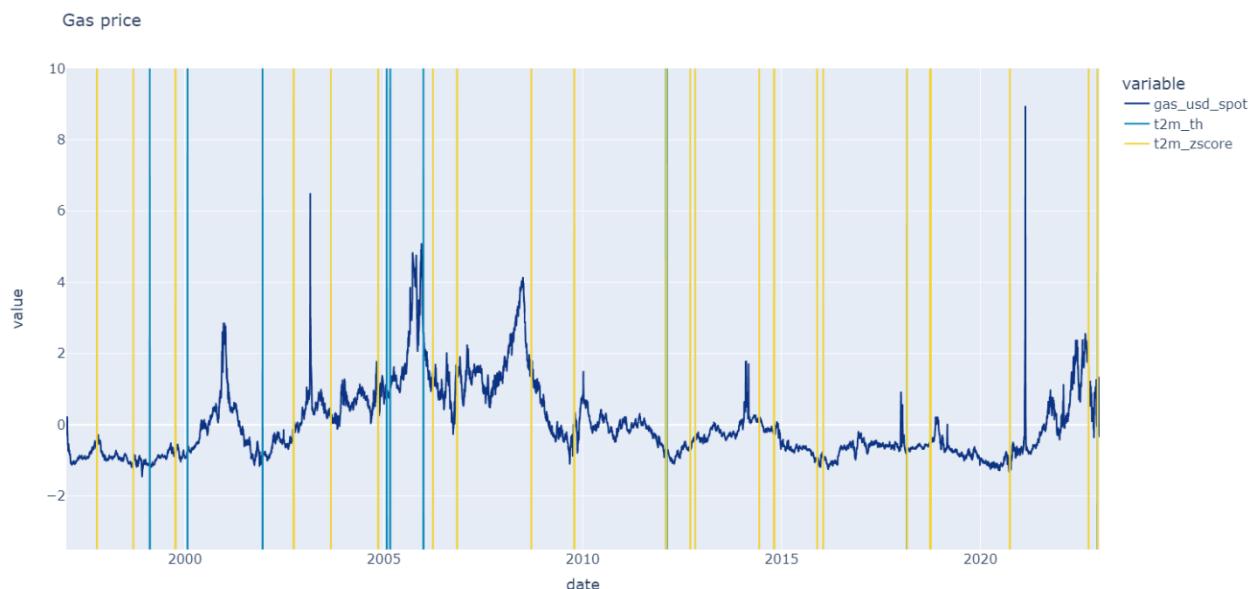


Figure 24 Gas price with z-score and threshold temperature peaks

## 5 Discussion

### 5.1 Polar Vortex Analysis

The conducted analysis of the polar vortex data, in combination with the outlined literature describing the impact on European weather, two partly successful indexes were developed and explored. Those two indexes propose one possibility of analysing and assessing the current state of the polar vortex over the defined region across the set pressure levels. Furthermore, the detection of the breakdown events only with the calculated indexes possess some skill, when compared against the SSW-events. Apart from the categorical breakdown event features, the current latitude or position as well as the metric at said position (temperature or windspeed) was successfully utilized as a feature for the machine learning models to predict the average temperature.

The impact on the machine learning model could only be precisely assessed for the RF classifier, as it possesses the feature importance of every metric. That being said, if the model is used as benchmark, the polar vortex features definitely hold an impact onto the model's performance.

Furthermore, a method of clustering the polar vortex data by utilizing the K-means clustering method was proposed in the form of image classification (unsupervised). Although further work would be needed to make it useable for further processing, it depicted a viable option of assessing the current state of the polar vortex.

### 5.2 Temperature Predictions and Machine Learning

Among the proposed and tested models, which include the RF-, MLP-, and RNN-models, the RF had the most accurate classification output. This was unexpected, when considering the given literature which mostly opted to use artificial neural networks for the task at hand. Possible explanations could be:

- The high complexity of the MLP and especially the RNN make it difficult to find the optimal model parameters and therefore optimize the model to its full potential.
- Although time series data was utilized, the use case of classifications might impact the outcome when compared to other models. All the reviewed literature did use regression models.
- The input parameters are more suited to an ensemble model, rather than a neural network.

Both, the RF- and MLP-classifiers were able to beat the set accuracy goal of 0.5 for the t1 and t2 predictions and the 0.25 for the overall accuracy. Both models would therefore be useable for the presented use case. Furthermore, it is interesting to note that MLP classifier possessed a higher accuracy on the t2 vector on the test set, which was not expected. The model accuracies on the test set, after thorough optimization and multiple iterations were:

Model	Accuracy t1	Accuracy t2	Accuracy
RFC	0.665	0.660	0.473
MLC	0.559	0.624	0.366

Figure 25 Prediction accuracies - Test set

The chosen weather indexes, as the feature importance of the RF model displayed, were quite significant for the performance or predictive skill. But the most important factor was the year as an integer value. This phenomenon was explained by the effects of steadily rising temperatures, which cause the average temperature to be higher as well, leading to more above categorised temperatures in recent years. It can be assumed, that the cause of this shift is global warming, although more research would be needed to make a conclusive statement.

The MLP regressor, showed, that comparable, but slightly worse performance was achieved with the given input parameters, when compared to the stated benchmark model. But because this was not the focus of the project, the model still would have some potential for improvements.

### **5.3 Gas Price and Outdoor Temperature Correlations**

Although the literature stated that average temperature in Europe has some effect on the gas prices, it was not possible to recreate and detect this impact for Switzerland specifically. Neither based on the normalized temperature, nor the classification of below and above average temperatures. The results are therefore not conclusive. It can be assumed, that if more resources and time would have been invested on this topic, a correlation or even causation could have emerged.

### **5.4 Additional Findings**

It was surprising to see, how much resources and time the Nu SVM did use to train a single model. The assumption was, that it would be more time efficient in training than the MLP and RNN. But this assumption turned out to be wrong. Two contributing factors were the size of the shifting window, which introduced a much higher dimensionality and the size of the data frame.

### **5.5 Limitations**

The main limitations which had an impact on the scope of the conducted experiments were:

- Time: Due to the time restriction, a lot of potential improvements and further steps could not be taken and had to be neglected to cover the whole scope of the thesis.
- Processing power: The trained neural networks and the shifting window turned out to be much more extensive than initially expected. Thus, leading to insufficiently available resources to train the RNNs. This in combination with the set time frame did not allow to optimize the RNNs to their full potential.
- Limited previous knowledge:
  - o The field of metrology is highly complex and has an enormous amount of available knowledge and research to compile and consider. That led to extensive and time intense research on certain topics to gain the needed competences to assess certain factors and indicators.
  - o Using models for image clustering, especially more complicated libraries like keras need a vast knowledge on how to apply the models correctly. Because the author only had limited exposure to this subject, the proposed image clustering method may be flawed.
- Changing requirements and methods: During the project some of the methods changed due to various reasons, which resulted in redownloading and recompiling data, as well as rewriting algorithms and indicators and retraining of models.

# 6 Conclusion

## 6.1 Summary

Two of the three goals which were set out could be achieved. A custom index to assess the current situation of the polar vortex as well as the impact on above and below average months was evaluated successfully. Furthermore, two machine learning models (RF, MLP) were able to beat the benchmark of a random guess on the classification of the historical temperature values for Switzerland. The direct correlation and causations of the average temperature on the gas spot prices could not be proven. Various factors may play into this, of which limited time was one of the main ones. In addition to the custom indexes, a method for applying image clustering to assess the polar vortex was proposed, but further research is needed.

This concludes that the proposed models can be used for long term temperature forecasting in any field of application. But the applications for energy trading remain in the focus, because of the rather imprecise nature of the models and the unusual way of the prediction.

## 6.2 Further Research

There are many possibilities of further research, which could be conducted to improve the overall results. The following listing may not be conclusive:

- Analysis of the polar vortex over a wider area, optimally over the whole polar region. Usually, the more data is available for a machine learning model, especially neural networks, the better the output is. But the provided data would have to be compiled in a useful manner.
- Implementing the created clusters of the polar vortex data into the machine learning models. If the clustering is done correctly, this would be another way to further enhance the machine learning models performance.
- Another approach which could be suggested for analysing the polar vortex is, to compile the given temperature data as a grey scale image (matrix), in the same way it was done for the wind speeds. The image data could then be used to create an unsupervised clustering model, creating categories or states of the polar vortex.
- Analysing and rewriting the custom indexes for detecting a polar vortex breakdown (or weak polar vortex) over one pressure level, with the longitude and latitude as x-and y-axis. This in combination with the before mentioned further step, could drastically improve the analysis of the polar vortex.
- As seen in the results of the RF model, the importance of past data points did not drop off sharply at the end of the 30-day shifting window. This suggests that a larger set of datapoints (e.g., a shifting window of 60 days) could be explored to further improve the models. The downside would be, that the training time would increase drastically, especially for the neural networks.
- The Nu SVC could be tested further. As it was mentioned in some papers, it could create a useable output.
- An interesting comparison could be proposed by optimizing a regression model and then manually transform its output to a categorical value (above or below average). With a manual layer in place which would handle the transformation, the classification metrics could be applied to the model and be compared to the classification models. It is possible the RNN would outperform the RF and MLP models with such a setup.

- After reaching a satisfying performance on the model, some interviews could be conducted with experts from the concerning industry. This could yield further information about the exact application and perhaps further possibilities for improvements.

## 7 List of Illustrations

Figure 1 Heating energy by source	5
Figure 2 ENSO [37]	11
Figure 3 MJO [38]	12
Figure 4 Polar vortex [12]	12
Figure 5 CRISP-DM [46]	15
Figure 6 General weather data area map	17
Figure 7 Confusion matrix schema [73]	24
Figure 8 SOI comparison - scaled values	27
Figure 10 Polar vortex - Data resolution	28
Figure 11 Correlation matrix	29
Figure 12 Temperature over time	30
Figure 13 Polar vortex breakdown event: left index 2, right index 4	33
Figure 14 PVT and PVS latitude comparison	34
Figure 15 Polar vortex breakdown and SSW comparison	35
Figure 16 Index latitude and temperature corelation	36
Figure 17 Index latitude and temperature corelation	36
Figure 18 t2m mean values	37
Figure 19 Breakdown event and category comparison	38
Figure 20 Random Forest confusion matrix	40
Figure 21 Feature importance by feature and time	41
Figure 22 Random Forest confusion matrix	45
Figure 23 Standardized gas price and t2m	48
Figure 24 Gas price and temperature correlations (scatter)	48
Figure 25 Gas price with z-score and threshold temperature peaks	49
Figure 26 Prediction accuracies - Test set	50
Figure 27 Temperature distribution (target vector)	60
Figure 28 SOI	61
Figure 29 MJO	61
Figure 30 AO	62
Figure 31 NAO	62
Figure 32 Moving average 30 days – SOI	63
Figure 33 Total polar vortex breakdown events	63
Figure 34 Same target vector categories distribution	64
Figure 35 Trend distribution of sets	64
Figure 36 Polar vortex wind speed and temperature	65
Figure 37 K-Means cluster distribution	66
Figure 38 Random Forest classifier fitting graph	66
Figure 39 Cluster distribution by month	67

## 8 List of Tables

Table 1 General weather data - API references	26
Table 2 Polar vortex attributes - API references	27
Table 3 Left to right: Main data frame, polar vortex data frame	28
Table 4 Index and applies parameters	34
Table 5 Random forest parameter tuning	39
Table 6 Optimized random forest accuracies	40
Table 7 Multi-layer perceptron parameter tuning	41
Table 8 Multi-layer perceptron accuracies	42
Table 9 Multi-layer perceptron confusion matrix	43
Table 10 Recurrent neural network parameters	43
Table 11 Recurrent neural network results	44
Table 12 Random forest classifier overfit parameters	45
Table 13 Random forest classifier performance (run 2)	45
Table 14 Multi-layer perceptron parameters	46
Table 15 Multi-layer perceptron results (unstandardized)	46
Table 16 Cluster description	47

## 9 List of Equations

$x = xN - 1$ [48]	17
$SOI = -1 * Pdiff - PdiffavSDPdiff + 2$ [49]	17
$SD = x - \mu 2N + 3$ [50]	17
$y = y1 + (x - x1)(y2 - y1)(x2 - x1) + 4$ [53]	18
$wind speed [m/s] = v = u2 + v2 + 5$ [57]	18
$wind direction [deg] = \tan^{-1} vu * 1802\pi + 6$ [57]	18
$r = n xy - x yn x2 - x2 n y2 - y2 + 7$ [58]	19
$year sine = \sin t * 2\pi n + 8$ [59]	19
$year cos = \cos t * 2\pi n + 9$ [59]	19
$ma = xt + xt - 1 + \dots + xt - n n + 10$ [60]	19
$yip = y1 + x - x1 y2 - y1 x2 - x1 + 11$ [61]	19
$lati1 = wsmean * 1 - wsstd - wsminstdwsmaxstd - wsminstd + 12$	20
$lati1 = wsmean * wws * 1 - wsstd - wsminstdwsmaxstd - wsminstd + wt * 1 - tstd - tminstdtmaxstd - tminstd + 13$	20
$lati2 = \max(d1, d2, \dots, dn) \geq dmean * th, latd \leq \max(d1, d2, \dots, dn) < dmean * th, 0 + 14$	20
$ind3 = 1 - lat - latminlatmax - latmin m - mminmmax - mmin - lat - latminlatmax - latminm - mminmmax - mmin + 15$ [62]	21
$lati4 = \max(m1, m2, \dots, mn) \geq mmean * th, latm \leq \max(m1, m2, \dots, mn) < mmean * th, 0 + 16$	21
$z = x - \mu train \sigma train + 17$ [63]	22
$f(x) = x < 0, 0 \leq x \geq 0, x + 18$ [69]	23
$f(x) = x < 0, a(ex - 1)x \geq 0, x + 19$ [71]	23
$acc = TP + TN / (TP + FP + TN + FN) + 20$ [73]	24
$R2 = yi - y2yi - y2 + 21$ [74]	24
$RMSE = \sqrt{\frac{1}{n} \sum (yi - \hat{yi})^2} + 22$ [74]	24

## 10 References

- [1] World Meteorological Organization, "Weather Forecasting," [Online]. Available: <https://youth.wmo.int/en/what-we-do/weather>. [Accessed 23 04 2023].
- [2] SciJinks, "How Reliable Are Weather Forecasts?," [Online]. Available: <https://scijinks.gov/forecast-reliability/>. [Accessed 23 04 2023].
- [3] E. McNamara, "The Bigger Picture | Energy Trading and the Insufficiency of Weather Forecasts," *EET&D Magazine*, vol. 23, no. q4, 4.
- [4] SWI Swiss Info, "Some 58% of Swiss buildings heated with fossil fuels," 06 10 2022. [Online]. Available: <https://www.swissinfo.ch/eng/business/some-58--of-swiss-buildings-heated-with-fossil-fuels/47958720>. [Accessed 23 04 2023].
- [5] Gaznat SA head office, "Natural gas - Supply," [Online]. Available: <https://www.gaznat.ch/en-39-supply.html>. [Accessed 23 04 2023].
- [6] World Economic Forum, "Russian energy imports to Europe are falling - here's what that means for the energy transition," 13 12 2022. [Online]. Available: <https://www.weforum.org/agenda/2022/12/europe-russia-energy-imports/>. [Accessed 25 04 2023].
- [7] K. Beckmann, "Energy Demand for Heating," ClimateXChange, Heriot Watt University, 2015.
- [8] Federal Office of Meteorology and Climatology MeteoSwiss, "ECMWF," [Online]. Available: <https://www.meteoswiss.admin.ch/about-us/research-and-cooperation/international-cooperation/ecmwf.html>. [Accessed 15 04 2023].
- [9] National centres for environmental information, "Global Forecast System (GFS)," [Online]. Available: <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>. [Accessed 12 04 2023].
- [10] National Center for Environmental Information, "Arctic Oscillation (AO)," 2023. [Online]. Available: <https://www.ncei.noaa.gov/access/monitoring/ao/>. [Accessed 15 04 2023].
- [11] National Center for Environmental Information, "North Atlantic Oscillation (NAO)," [Online]. Available: <https://www.ncei.noaa.gov/access/monitoring/nao/>. [Accessed 19 04 2023].
- [12] P. Duginski, "If a warm U.S. winter was 'a preview of global warming,' what part did a polar vortex play?," Los Angeles Times, 28 03 2020. [Online]. Available: <https://www.latimes.com/california/story/2020-03-28/if-a-warm-u-s-winter-was-a-preview-of-global-warming-what-part-did-a-polar-vortex-play>. [Accessed 15 04 2023].
- [13] R. Lindsey, "Understanding the Arctic polar vortex," Climate Gov, 05 03 2021. [Online]. Available: <https://www.climate.gov/news-features/understanding-climate/understanding-arctic-polar-vortex>. [Accessed 15 04 2023].
- [14] D. B. Halim Tatli, "The Effects of Climate and Oil Prices on Residential Natural Gas Prices: With an Application to 11 OECD Countries," *Acta Montanistica Slovaca*, vol. 27, no. 3, pp. 710-722, 2022.
- [15] M. H. Pesaran, "On the Interpretation of Panel Unit Root Tests," University of Cambridge and USC, Trinity College, Cambridge, CB2 1TQ, England, 2011.
- [16] Q. J. Y. F. Jiang-Bo Geng, "The behaviour mechanism analysis of regional natural gas prices: A multi-scale perspective," *Energy*, vol. 101, no. 1, pp. 266-277, 2016.
- [17] B. Arrington, "Which Weather Forecasting Model is the Best?," Captain's Tips, 2016. [Online]. Available: <https://www.marinalife.com/articles/weather-forecasting-model-best-know-go>. [Accessed 12 04 2023].
- [18] weather.us, "Model charts," [Online]. Available: <https://weather.us/model-charts>. [Accessed 12 04 2023].
- [19] Meteologix, "Model charts," [Online]. Available: <https://meteologix.com/jo/model-charts>. [Accessed 12 04 2023].
- [20] windy.app, "What is GFS weather model and how it works," [Online]. Available: <https://windy.app/blog/what-is-gfs-weather-forecast-model.html>. [Accessed 12 04 2023].

- [21] Wikipedia, "Global Forecast System," 23 01 2023. [Online]. Available: [https://en.wikipedia.org/wiki/Global\\_Forecast\\_System](https://en.wikipedia.org/wiki/Global_Forecast_System). [Accessed 12 04 2023].
- [22] European Centre for Medium-Range Weather Forecasts, "About our forecasts," [Online]. Available: <https://www.ecmwf.int/en/forecasts/documentation-and-support>. [Accessed 15 04 2023].
- [23] Copernicus, "Long-term Monthly forecast of temperature and rainfall anomalies," [Online]. Available: <https://effis.jrc.ec.europa.eu/apps/effis.longterm.forecasts/>. [Accessed 15 04 2023].
- [24] WINDY.APP, "What is ICON weather model and how it works," [Online]. Available: <https://windy.app/blog/what-is-icon-weather-model-forecast.html>. [Accessed 15 04 2023].
- [25] J.-P. Lang, "ICON :: Icosahedral Nonhydrostatic Weather and Climate Model," Mpimet, [Online]. Available: <https://code.mpimet.mpg.de/projects/iconpublic>. [Accessed 15 04 2023].
- [26] G. M. A. B. J. R. Jenny Cifuentes, "Air Temperature Forecasting Using Machine Learning Techniques: A Review," *Energies*, vol. 13, no. 16, 2020.
- [27] Z. M. Mohsen Hayati, "Application of Artificial Neural Networks for," *World Academy of Science, Engineering and Technology*, vol. 28, pp. 275-279, 2007.
- [28] R. Abdel-Aal, "Hourly temperature forecasting using abductive networks," *Engineering Applications of Artificial Intelligence*, vol. 17, pp. 543-556, 2004.
- [29] M. Y. Radhika, "Atmospheric Temperature Prediction using," *International Journal of Computer Theory and Engineering*, vol. 1, no. 1, pp. 55-58, 2009.
- [30] P. M. B. & B. S. Aghelpour, "Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA," *Theor Appl Climatol*, vol. 138, p. 1471-1480, 2019.
- [31] R. S. P. R. E. C. S. Z. Y. L. R. C. Toni Toharudin, "Employing long short-term memory and Facebook prophet model in air temperature forecasting," *Communications in Statistics - Simulation and Computation*, vol. 52, no. 2, pp. 279-290, 2021.
- [32] J. P.-A. C. P.-R. J. D. S. S. S.-S. D. Fister, "Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques," *Applied Soft Computing*, vol. 136, 2023.
- [33] S. M. B. S. J. K. H. V. Trang Thi Kieu Tran, "A Review of Neural Networks for Air Temperature Forecasting," *Water*, vol. 13, no. 9, 2021.
- [34] E. S. S. R. G. S. D. Hossein Hassani, "Predicting global temperature anomaly: A definitive investigation using an ensemble of twelve competing forecasting models," *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 121-139, 2018.
- [35] "An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies," *Journal of Climate*, vol. 26, pp. 5260-5269, 2012.
- [36] M. T. A. A. S. a. I. H. A. A. S Wibisono, "Multivariate weather anomaly detection using DBSCAN clustering algorithm," *Journal of Physics: Conference Series*, vol. 1869, 2021.
- [37] Climate gov, "El Niño & La Niña (El Niño-Southern Oscillation)," 09 03 2022. [Online]. Available: <https://www.climate.gov/enso>. [Accessed 15 04 2023].
- [38] J. Gottschalck, "What is the MJO, and why do we care?," 31 12 2014. [Online]. Available: <https://www.climate.gov/news-features/blogs/enso/what-mjo-and-why-do-we-care>. [Accessed 15 04 2023].
- [39] A. Flis, "A strong Blocking system is forming over Greenland, changing the Weather patterns across Northern Hemisphere as we head into Spring," Severe Weather Europe, 05 03 2023. [Online]. Available: <https://www.severe-weather.eu/global-weather/strong-blocking-system-polar-vortex-collapse-cold-weather-united-states-canada-europe-fa/>. [Accessed 15 04 2023].
- [40] A. A. Scaife, "Impact of ENSO on European Climate," Met Office Hadley Centre, Exeter, UK, 2010.
- [41] C. C. School, "ENSO Essentials," [Online]. Available: <https://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/>. [Accessed 15 04 2023].
- [42] F. M. a. T. J. Frédéric Vitart, "Prediction of the Madden-Julian Oscillation and its impact on the European weather in the ECMWF monthly forecasts," ECMWF , RG2 9AX, United Kingdom, 2010.

- [43] Australian Government, "Madden-Julian Oscillation (MJO)," [Online]. Available: <http://www.bom.gov.au/climate/mjo/#tabs=Phase>. [Accessed 15 04 2023].
- [44] NASA Ozone Watch, "What is a Polar Stratospheric Warming?," NASA, 14 02 2023. [Online]. Available: [https://ozonewatch.gsfc.nasa.gov/facts/warming\\_NH.html](https://ozonewatch.gsfc.nasa.gov/facts/warming_NH.html). [Accessed 20 05 2023].
- [45] C. S. L. (. A. Butler), "Table of major mid-winter SSWs in reanalyses products," 02 2020. [Online]. Available: <https://csl.noaa.gov/groups/csl8/sswcompendium/majorevents.html>. [Accessed 20 05 2023].
- [46] N. Hotz, "What is CRISP DM?," Data Science Process Alliance, 19 01 2023. [Online]. Available: <https://www.datascience-pm.com/crisp-dm-2/>. [Accessed 15 04 2023].
- [47] "ERA5 hourly data on single levels from 1940 to present," Copernicus, [Online]. Available: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>. [Accessed 19 04 2023].
- [48] Biology for Life, "Measures of Central Tendency," [Online]. Available: <https://www.biologyforlife.com/measures-of-central-tendency.html>. [Accessed 16 04 2023].
- [49] Commonwealth of Australia , Bureau of Meteorology, "Southern Oscillation Index," [Online]. Available: <http://www.bom.gov.au/climate/glossary/soi.shtml>. [Accessed 25 04 2023].
- [50] Khan Academy, "Calculating standard deviation step by step," [Online]. Available: <https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/variance-standard-deviation-population/a/calculating-standard-deviation-step-by-step>. [Accessed 25 04 2023].
- [51] Australian Government, Bureau of Meteorology, "Madden-Julian Oscillation (MJO)," [Online]. Available: <http://www.bom.gov.au/climate/mjo/>. [Accessed 19 04 2023].
- [52] National Weather Service NAOO, "Arctic Oscillation," [Online]. Available: [https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily\\_ao\\_index/ao.shtml](https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.shtml). [Accessed 19 04 2023].
- [53] Cuemath, "Linear Interpolation Formula," [Online]. Available: <https://www.cuemath.com/linear-interpolation-formula/>. [Accessed 25 04 2023].
- [54] Copernicus, "ERA5 hourly data on pressure levels from 1940 to present," [Online]. Available: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form>. [Accessed 19 04 2023].
- [55] meteoblue, "Air pressure," [Online]. Available: <https://content.meteoblue.com/en/research-education/specifications/weather-variables/pressure>. [Accessed 07 05 2023].
- [56] firenzemeteo, "10 hPa temperature and geopotential," [Online]. Available: <https://www.firenzemeteo.it/en/teleconnections/10-hPa-temperature-geopotential.php>. [Accessed 07 05 2023].
- [57] Varsity Tutors, "Magnitude and Direction of Vectors," [Online]. Available: [https://www.varsitytutors.com/hotmath/hotmath\\_help/topics/magnitude-and-direction-of-vectors](https://www.varsitytutors.com/hotmath/hotmath_help/topics/magnitude-and-direction-of-vectors). [Accessed 26 04 2023].
- [58] S. Turney, "Pearson Correlation Coefficient (r) | Guide & Examples," 12 05 2022. [Online]. Available: <https://www.scribbr.com/statistics/pearson-correlation-coefficient/>. [Accessed 29 04 2023].
- [59] P.-L. Bescond, "Cyclical features encoding, it's about time!," Towards Data Science, 08 06 2020. [Online]. Available: <https://towardsdatascience.com/cyclical-features-encoding-its-about-time-ce23581845ca>. [Accessed 07 05 2023].
- [60] J. Fernando, "Moving Average (MA): Purpose, Uses, Formula, and Examples," Investopedia, 31 03 2023. [Online]. Available: <https://www.investopedia.com/terms/m/movingaverage.asp>. [Accessed 07 05 2023].
- [61] Cue Math, "Linear Interpolation Formula," [Online]. Available: <https://www.cuemath.com/linear-interpolation-formula/>. [Accessed 07 05 2023].
- [62] S. M. Cendrero, "Warning About Normalizing Data," Telefonica, 01 11 2018. [Online]. Available: <https://business.blogthinkbig.com/warning-about-normalizing-data/>. [Accessed 07 05 2023].

- [63] Statistics How To , “Standardized Values,” [Online]. Available: <https://www.statisticshowto.com/standardized-values-examples/>. [Accessed 17 05 2023].
- [64] T. Yiu, “Understanding Random Forest,” Towards Data Science, 12 06 2019. [Online]. Available: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>. [Accessed 17 05 2023].
- [65] scikit learn, “Support Vector Machines,” [Online]. Available: <https://scikit-learn.org/stable/modules/svm.html>. [Accessed 17 05 2023].
- [66] Ç. N. Akkaya Berke, Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases, 2019.
- [67] J. Brownlee, “How to Develop Multilayer Perceptron Models for Time Series Forecasting,” Machine Learning Mastery, 09 11 2018. [Online]. Available: <https://bobrupakroy.medium.com/multi-step-lstm-time-series-forecasting-6d9e4777883b>. [Accessed 17 05 2023].
- [68] H. Onnen, “Temporal Loops: Intro to Recurrent Neural Networks for Time Series Forecasting in Python,” Towards Data Science, 31 10 2021. [Online]. Available: <https://towardsdatascience.com/temporal-loops-intro-to-recurrent-neural-networks-for-time-series-forecasting-in-python-b0398963dc1f>. [Accessed 17 05 2023].
- [69] Z. Brodtman, “The Importance and Reasoning behind Activation Functions,” Towards Data Science, 15 11 2021. [Online]. Available: <https://towardsdatascience.com/the-importance-and-reasoning-behind-activation-functions-4dc00e74db41>. [Accessed 17 05 2023].
- [70] T. Böhm, “A first Introduction to SELUs and why you should start using them as your Activation Functions,” Towards Data Science, 28 08 2018. [Online]. Available: <https://towardsdatascience.com/gentle-introduction-to-selus-b19943068cd9>. [Accessed 17 05 2023].
- [71] T. Böhm, “A first Introduction to SELUs and why you should start using them as your Activation Functions,” Towards Data Science, 28 08 2018. [Online]. Available: <https://towardsdatascience.com/gentle-introduction-to-selus-b19943068cd9>. [Accessed 30 03 2023].
- [72] N. Sharma, “K-Means Clustering Explained,” MLOps Blog, 19 04 2023. [Online]. Available: <https://neptune.ai/blog/k-means-clustering>. [Accessed 10 06 2023].
- [73] S. Tran, “6 Useful Metrics to Evaluate Binary Classification Models,” The digital Skye, 19 04 2021. [Online]. Available: <https://thedigitalskye.com/2021/04/19/6-useful-metrics-to-evaluate-binary-classification-models/>. [Accessed 28 05 2023].
- [74] A. Kumar, “Mean Squared Error or R-Squared – Which one to use?,” Data Analytics, 04 04 2023. [Online]. Available: <https://vitalflux.com/mean-square-error-r-squared-which-one-to-use/>. [Accessed 03 06 2023].
- [75] J. Brakel, “Robust peak detection algorithm using z-scores,” Stack Overflow, 08 11 2020. [Online]. Available: <https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data/22640362#22640362>. [Accessed 29 05 2023].
- [76] National Weather Service NOAA, “Monthly Atmospheric and SST Indices,” [Online]. Available: <https://www.cpc.ncep.noaa.gov/data/indices/soi>. [Accessed 26 04 2023].
- [77] U.S. Energy Information Administration, “Natural Gas,” [Online]. Available: <https://www.eia.gov/dnav/ng/hist/rngwhhdD.htm>. [Accessed 30 05 2023].
- [78] National data buoy Center, “Are wind speeds the same over land as they are over the ocean?,” [Online]. Available: [https://www.ndbc.noaa.gov/education/windspeed\\_ans.shtml](https://www.ndbc.noaa.gov/education/windspeed_ans.shtml). [Accessed 30 04 2023].
- [79] Z. Brodtman, “The Importance and Reasoning behind Activation Functions,” Towards Data Science, 15 11 2021. [Online]. Available: <https://towardsdatascience.com/the-importance-and-reasoning-behind-activation-functions-4dc00e74db41>. [Accessed 30 03 2023].

# 11 Appendix

## 11.1 Thesis to jupyter notebook Chapter References

Title / Description	Thesis result chapter	Notebook chapter
Data understanding – Data gathering	4.1	0
Data understanding – Data exploration	4.2	1
Data preparation	4.3, 4.4, 1	2, 3
Classification - Modelling (RF)	4.4.2	3
Classification - Modelling (SVM)	4.4.3	4
Classification - Modelling (MLP)	4.4.4	5
Classification - Modelling (RNN)	4.4.5	6
Classification - Model optimization (RF)	4.5	3
Regression - Modelling (MLP)	4.6	5
Clustering - Modelling (K-Means)	4.7	8
Gas price analysis	4.8	7

## 11.2 Referenced Plots and Graphs

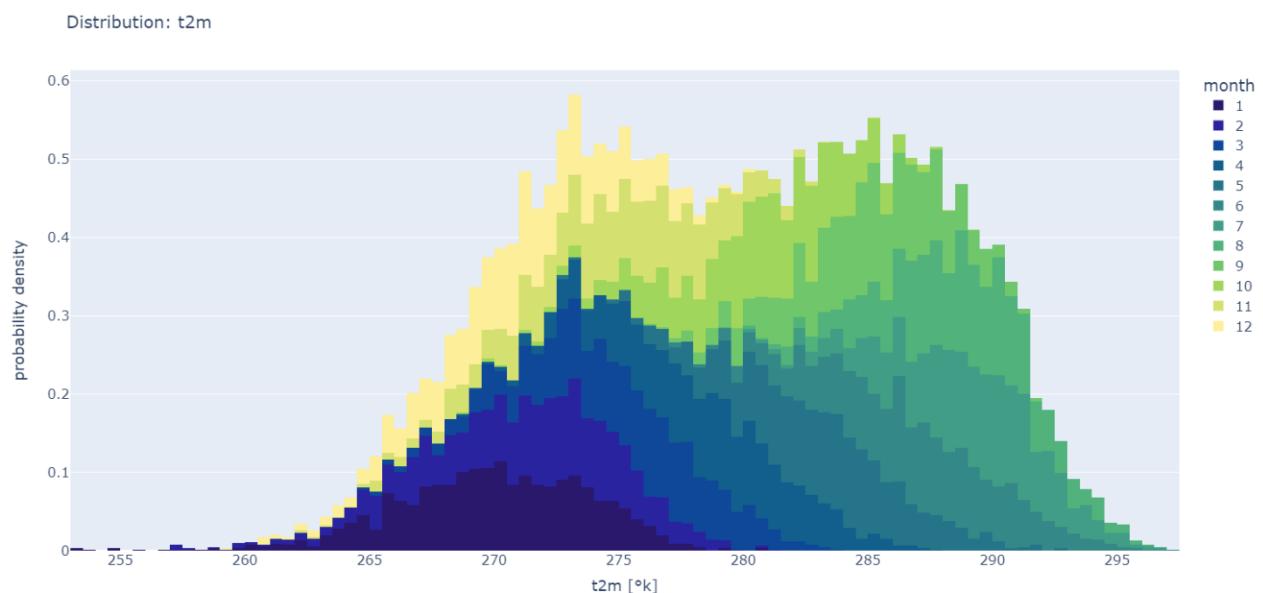


Figure 26 Temperature distribution (target vector)

SOI

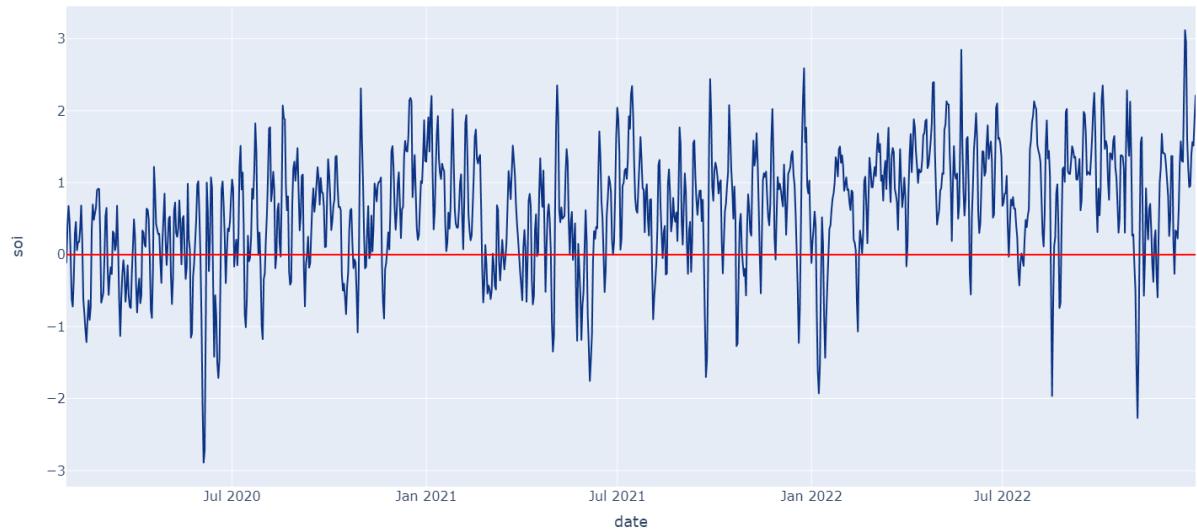


Figure 27 SOI

MJO amplitude

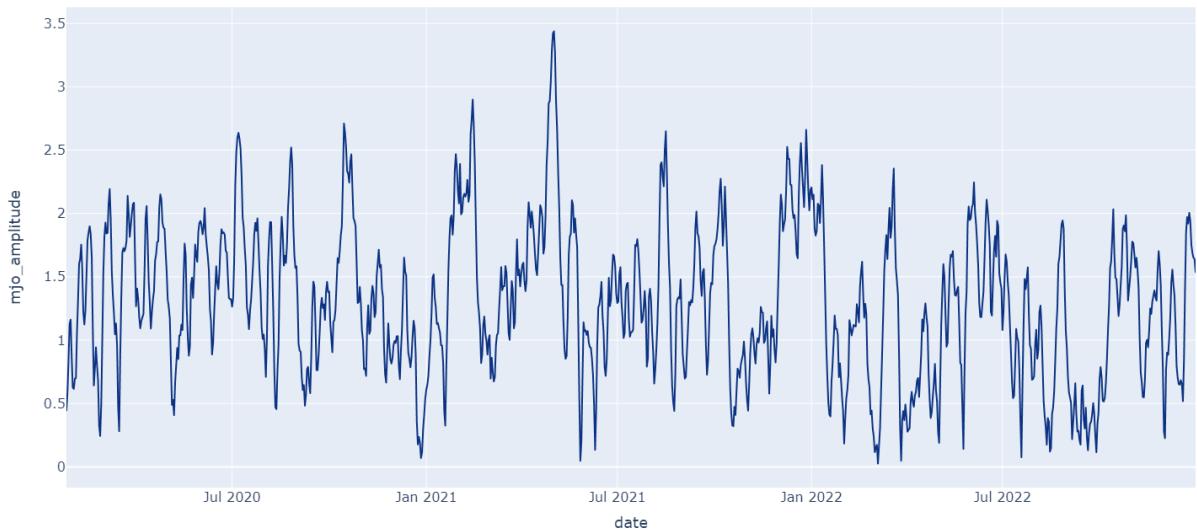


Figure 28 MJO



Figure 29 AO

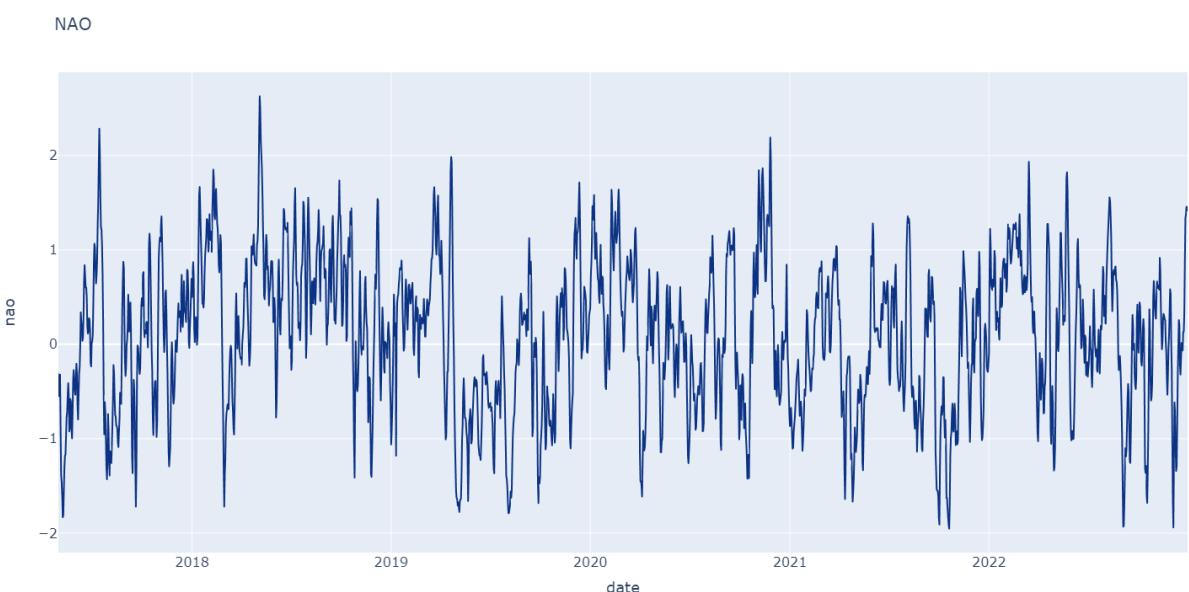


Figure 30 NAO

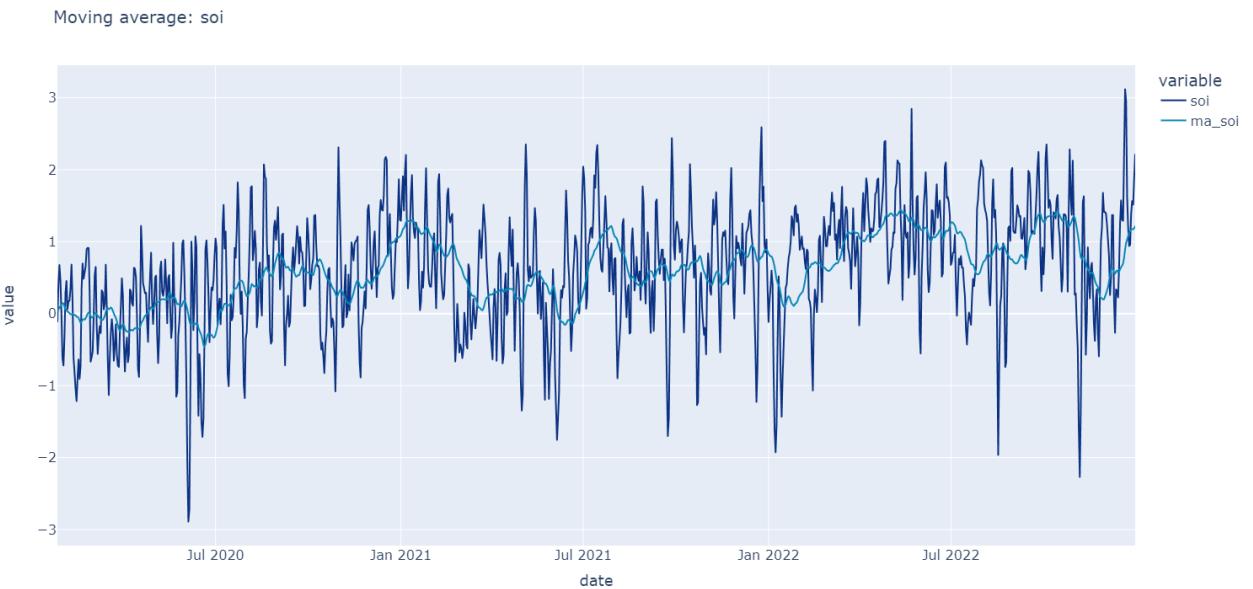


Figure 31 Moving average 30 days – SOI

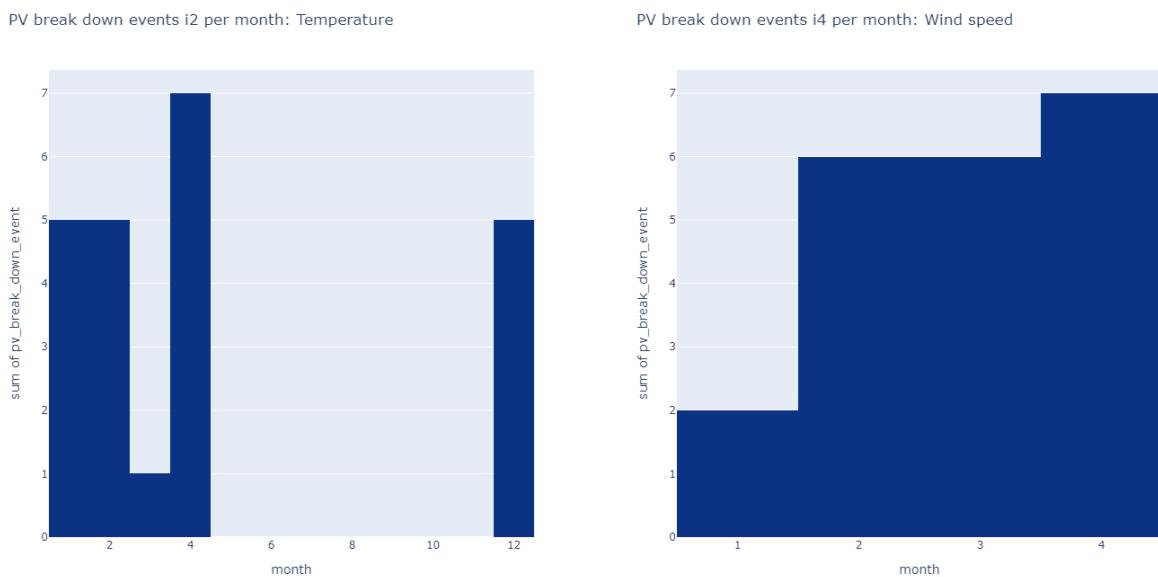


Figure 32 Total polar vortex breakdown events

Same category in target vector ( $t_1 = t_2$ )

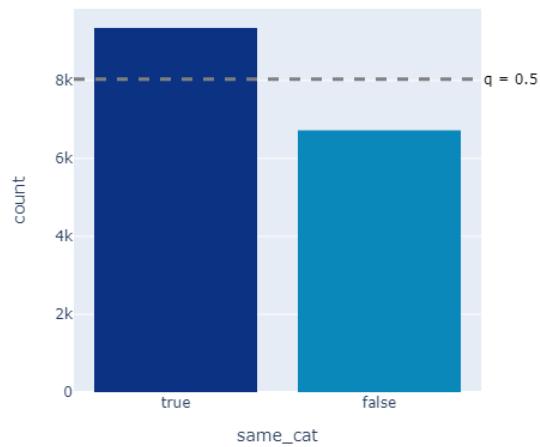


Figure 33 Same target vector categories distribution

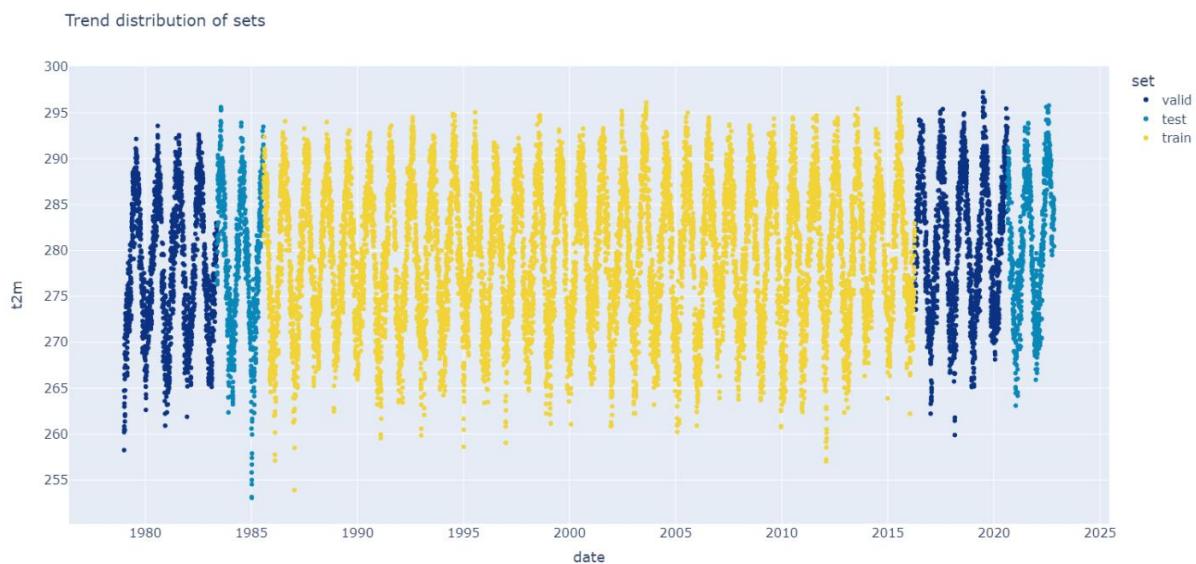
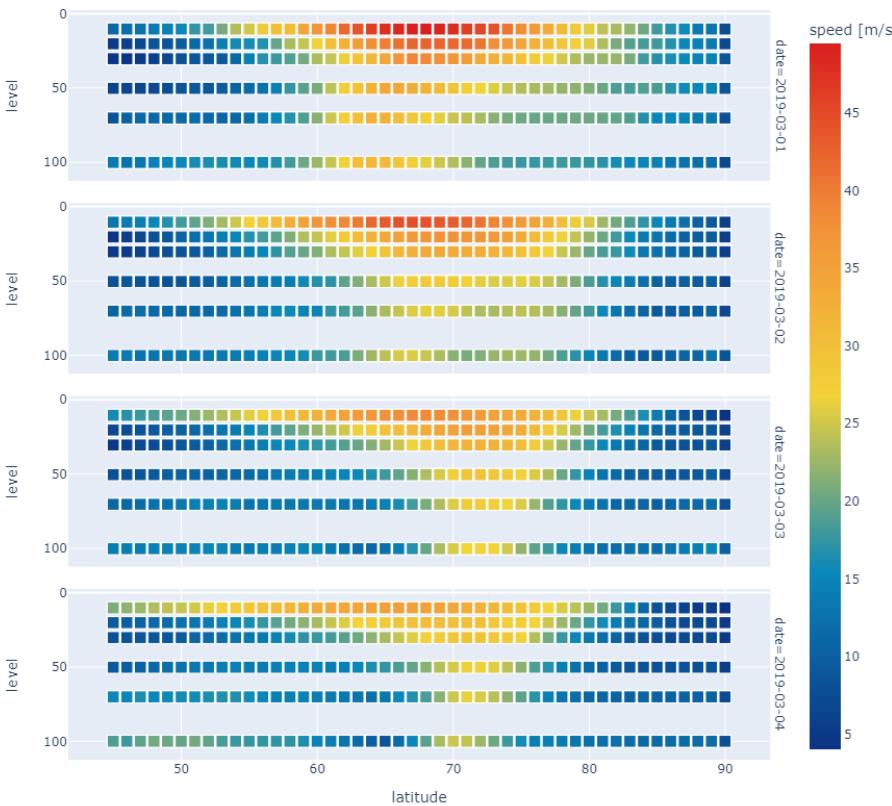


Figure 34 Trend distribution of sets

Polar vortex wind speed



Polar vortex temperature

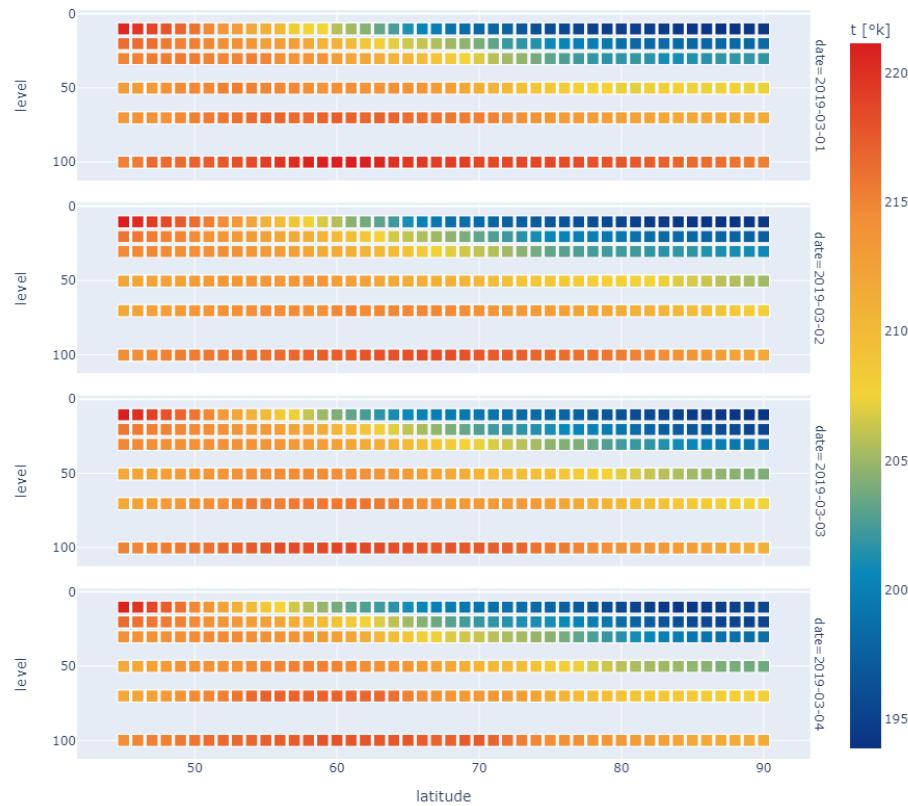


Figure 35 Polar vortex wind speed and temperature

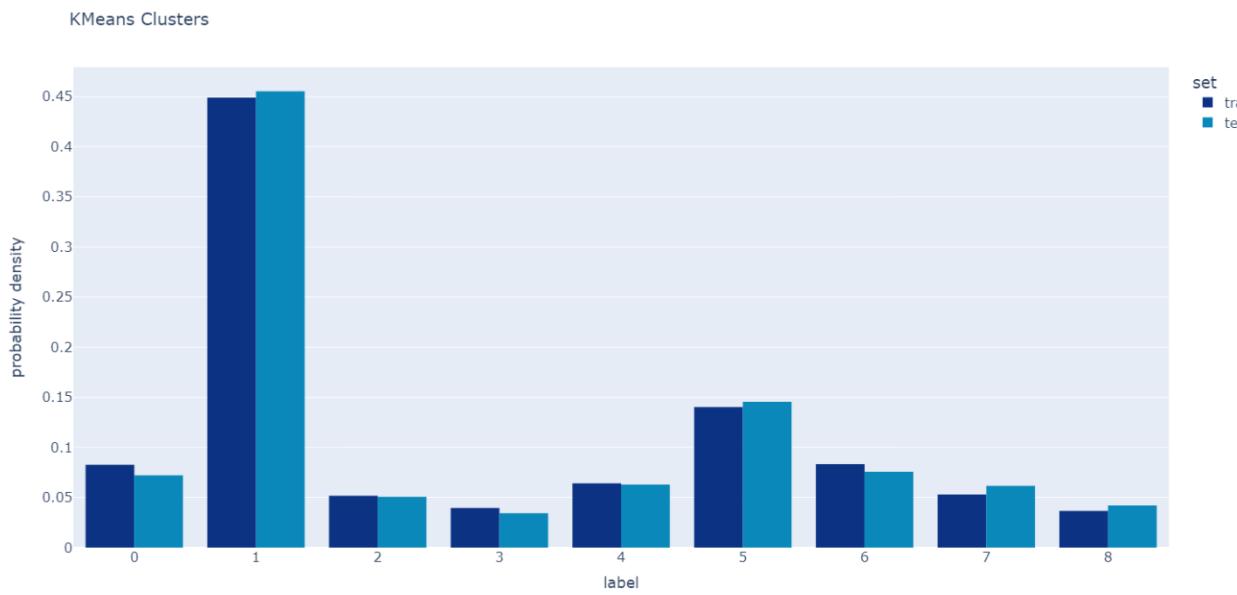


Figure 36 K-Means cluster distribution

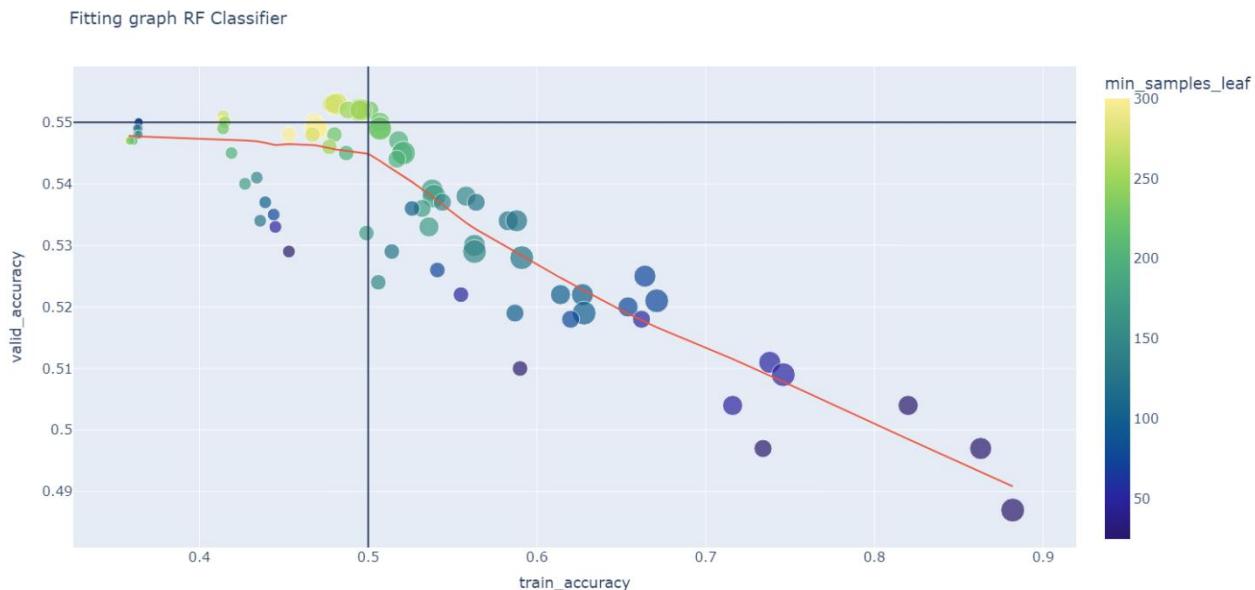


Figure 37 Random Forest classifier fitting graph

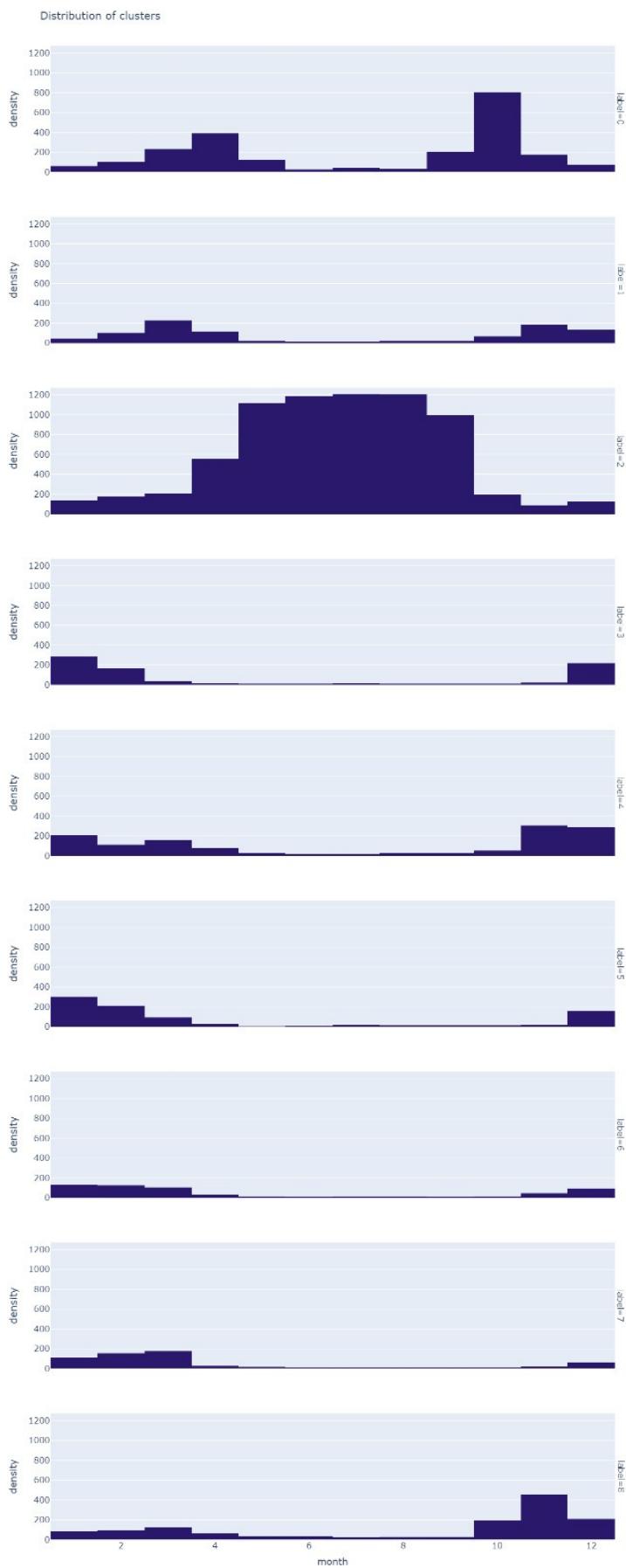
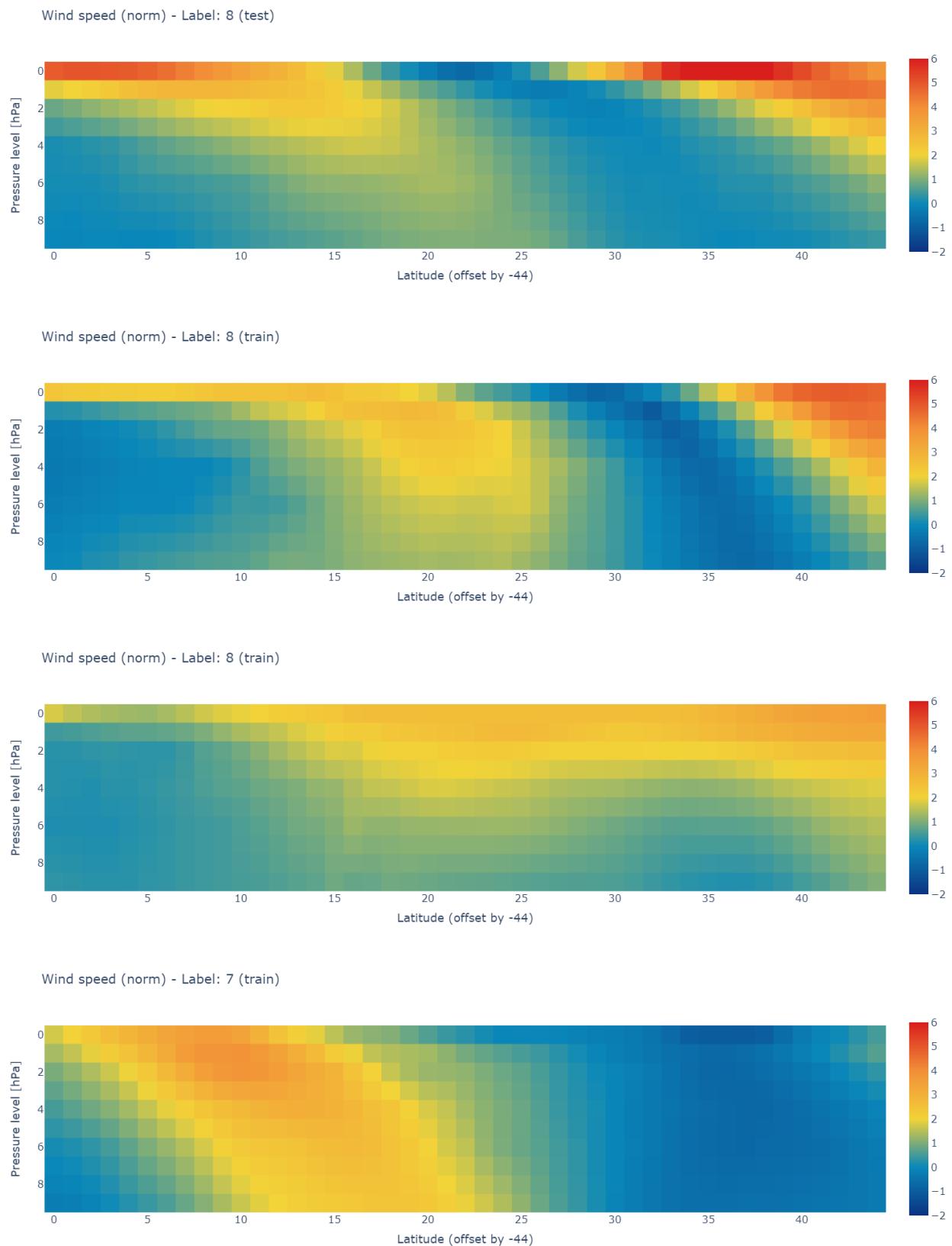
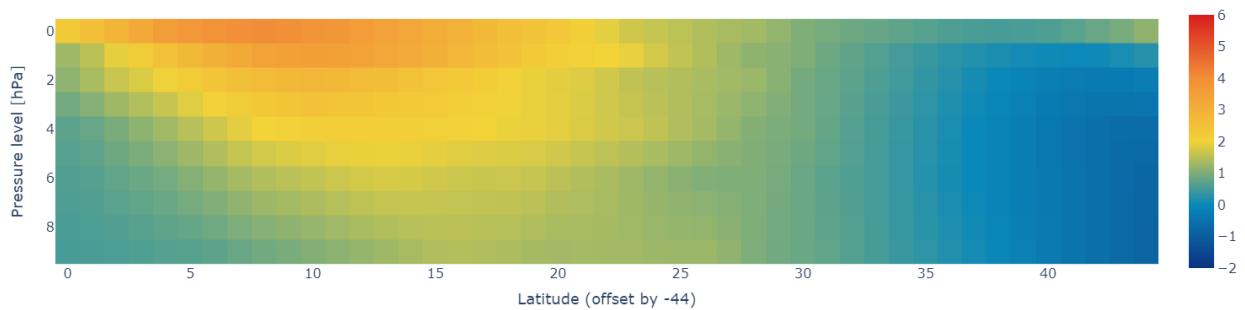


Figure 38 Cluster distribution by month

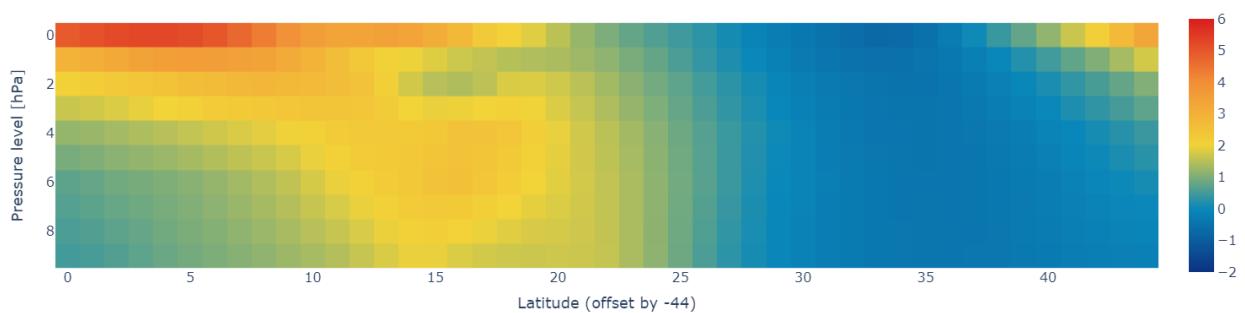
### 11.3 Clustering Sample



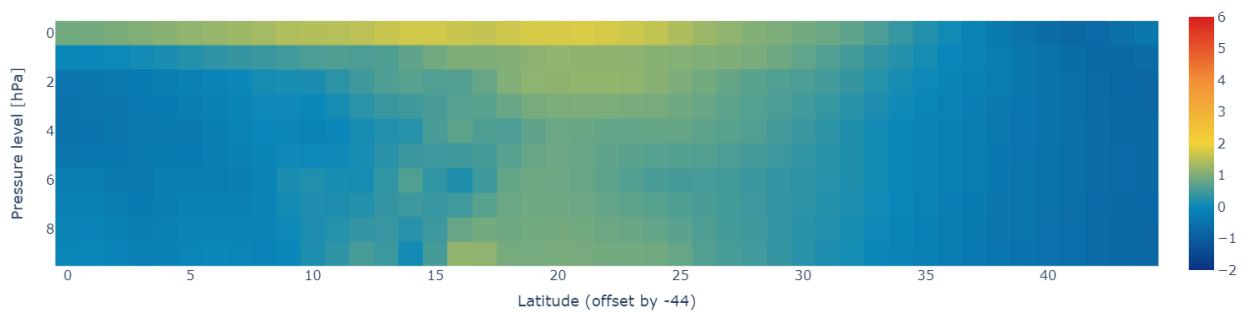
Wind speed (norm) - Label: 7 (train)



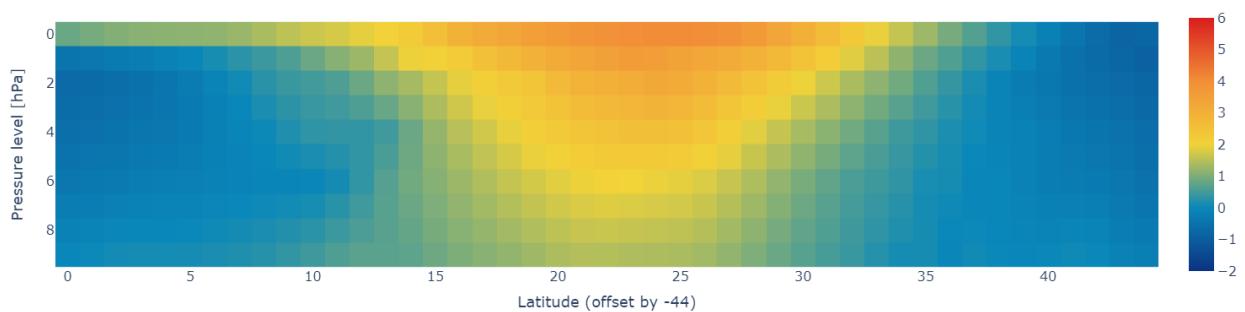
Wind speed (norm) - Label: 7 (train)



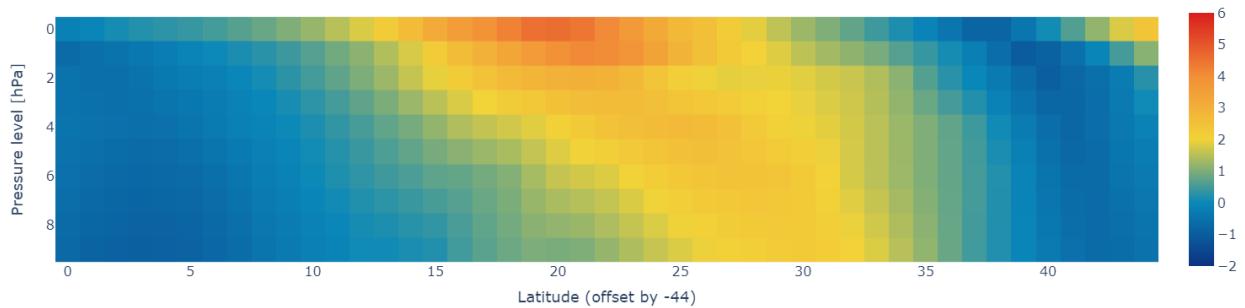
Wind speed (norm) - Label: 6 (train)



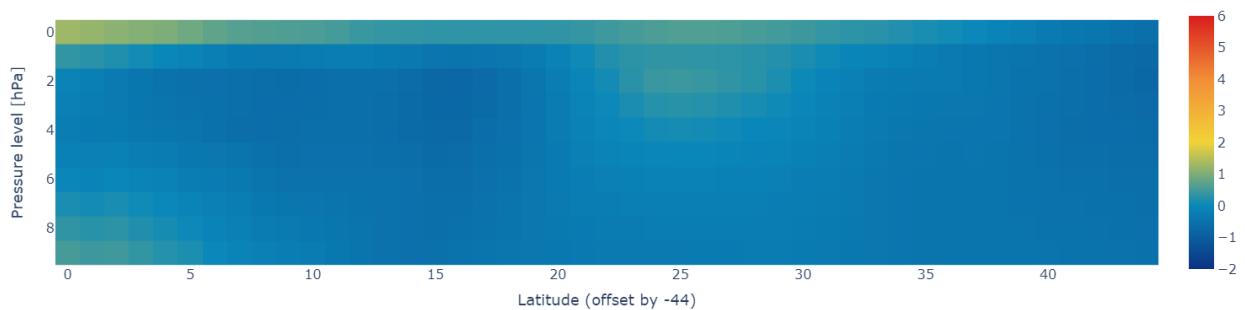
Wind speed (norm) - Label: 6 (train)



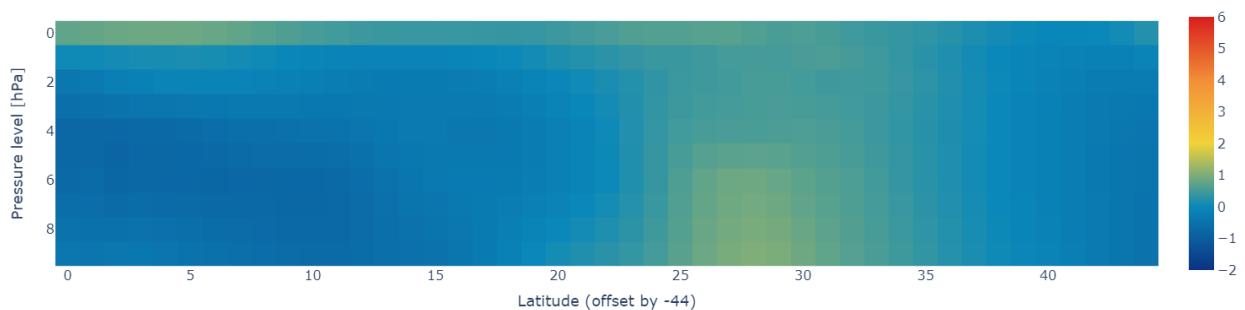
Wind speed (norm) - Label: 6 (train)



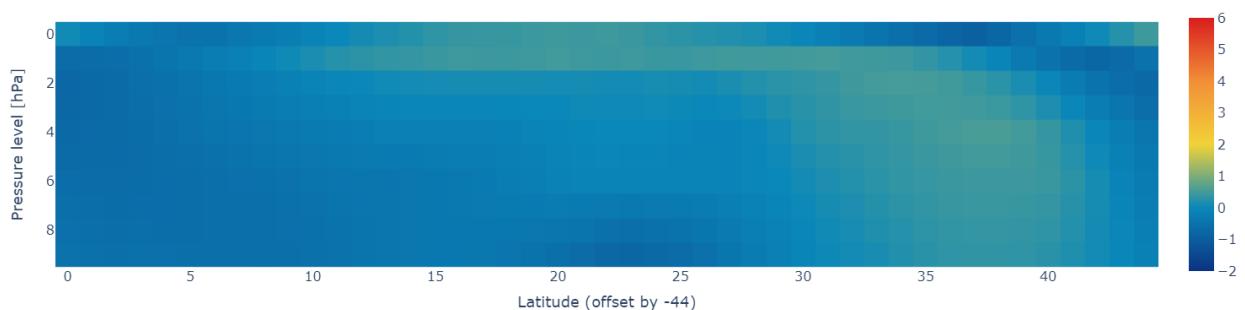
Wind speed (norm) - Label: 5 (train)



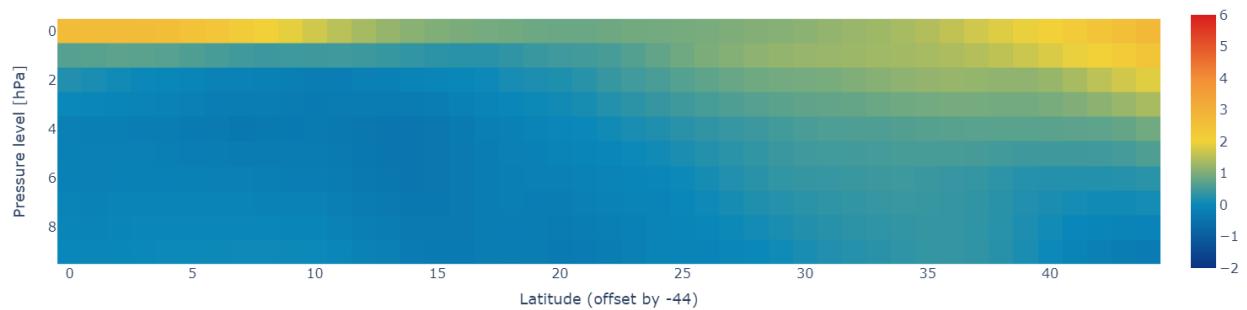
Wind speed (norm) - Label: 5 (test)



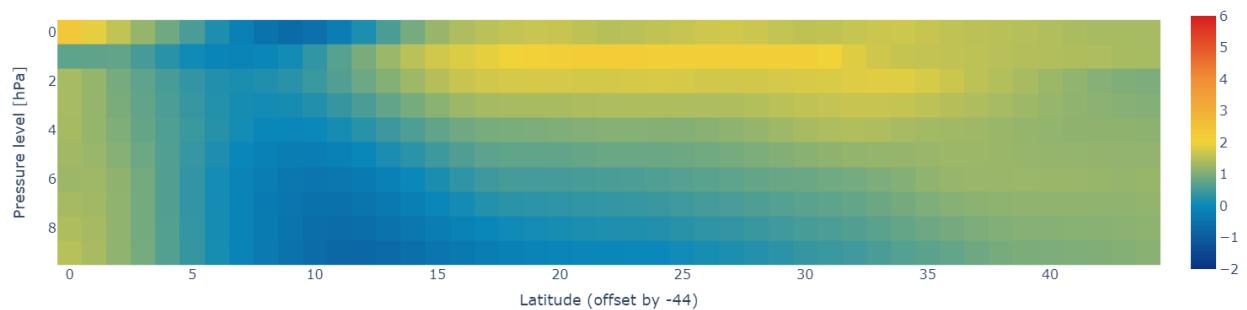
Wind speed (norm) - Label: 5 (train)



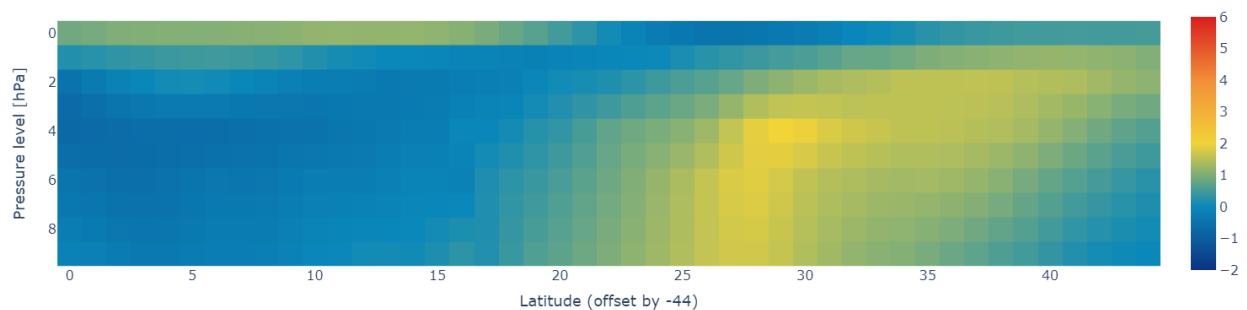
Wind speed (norm) - Label: 4 (train)



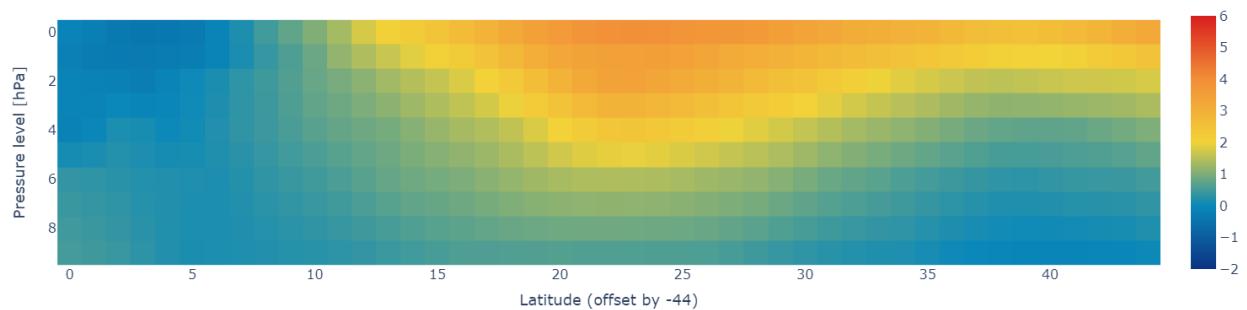
Wind speed (norm) - Label: 4 (train)



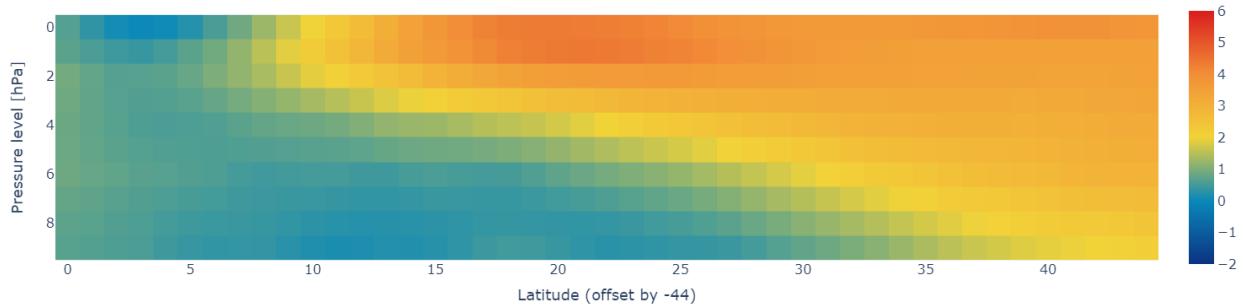
Wind speed (norm) - Label: 4 (train)



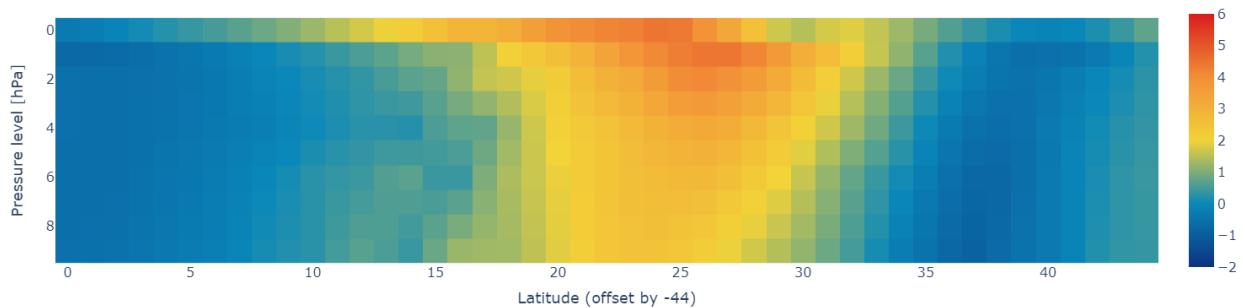
Wind speed (norm) - Label: 3 (train)



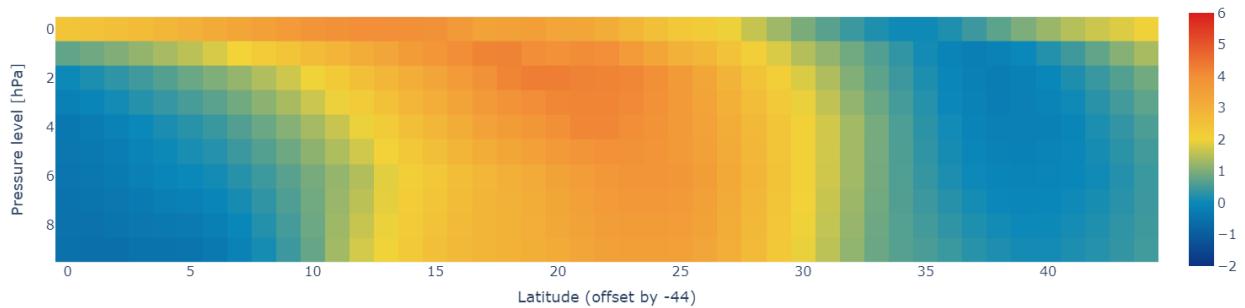
Wind speed (norm) - Label: 3 (train)



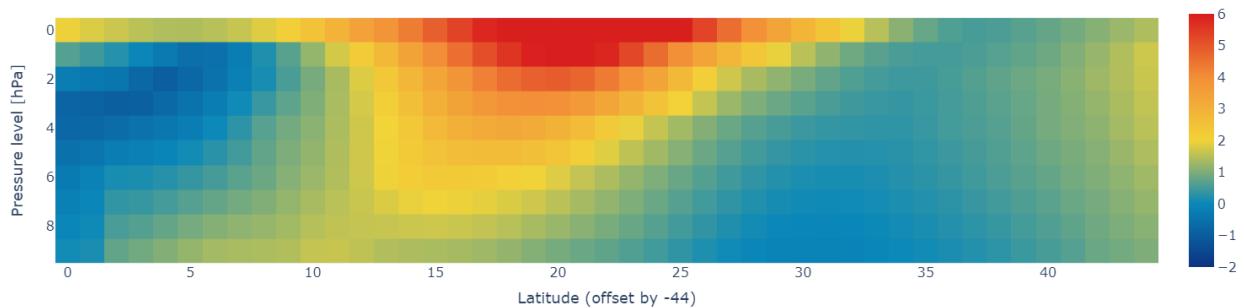
Wind speed (norm) - Label: 3 (train)



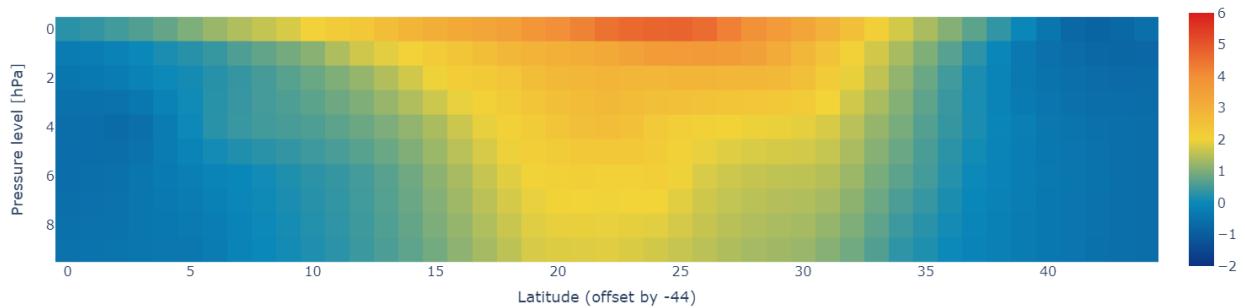
Wind speed (norm) - Label: 2 (train)



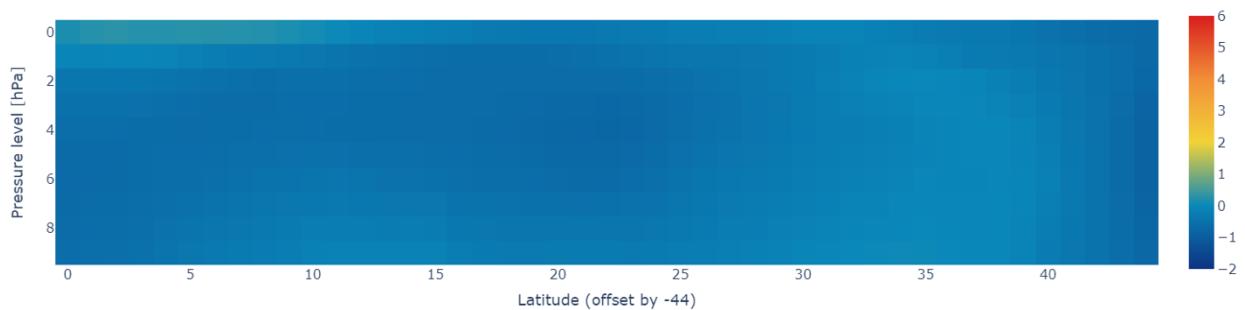
Wind speed (norm) - Label: 2 (test)



Wind speed (norm) - Label: 2 (test)



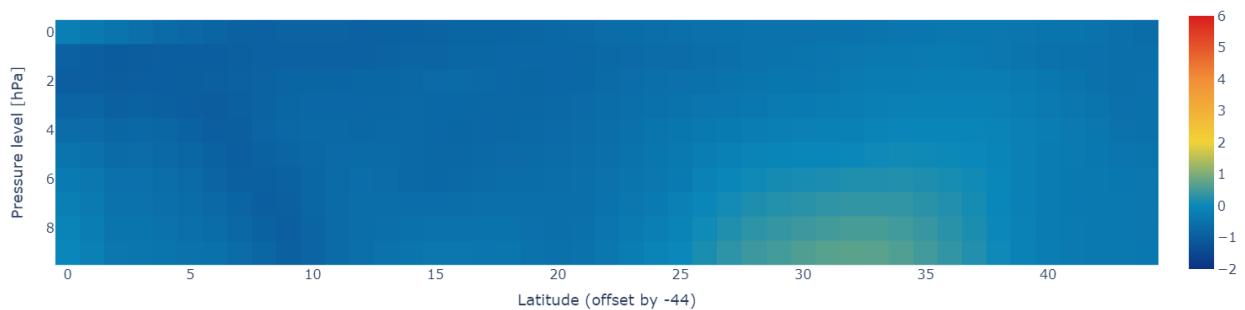
Wind speed (norm) - Label: 1 (train)



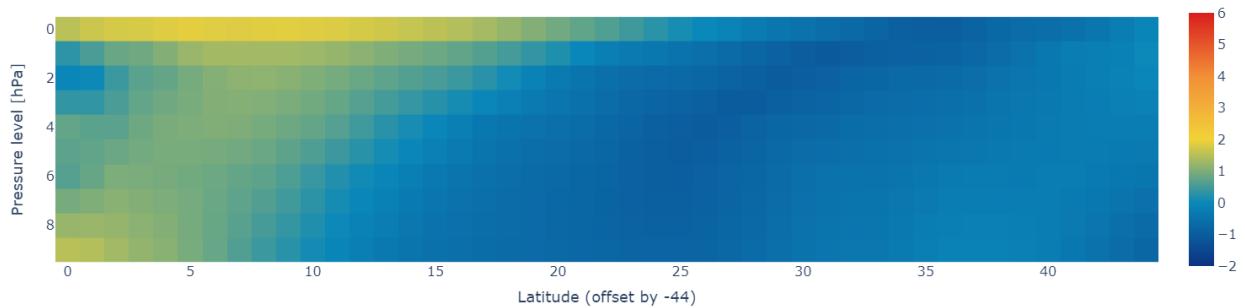
Wind speed (norm) - Label: 1 (train)



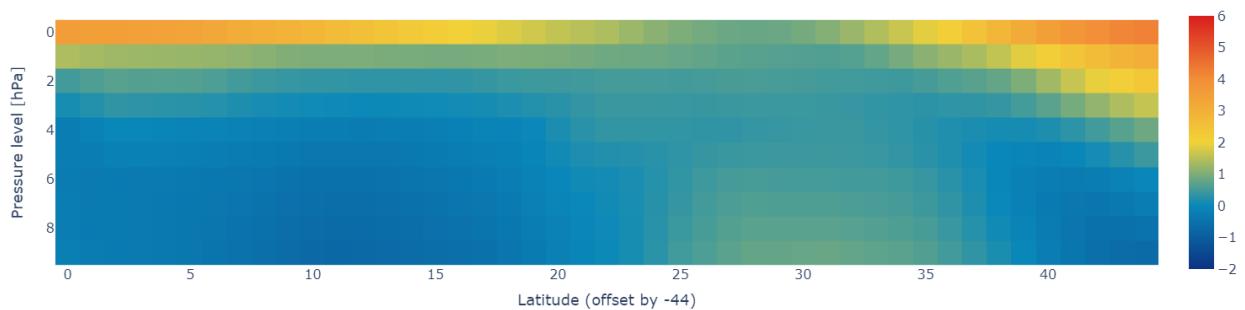
Wind speed (norm) - Label: 1 (test)



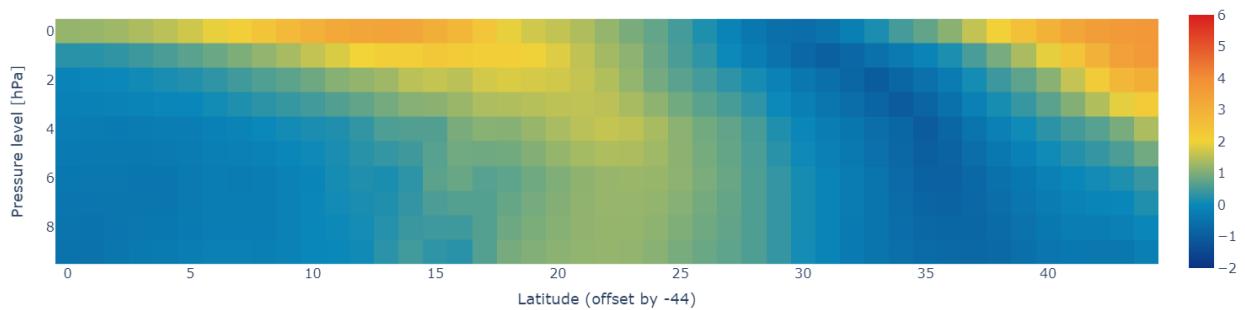
Wind speed (norm) - Label: 0 (train)



Wind speed (norm) - Label: 0 (train)

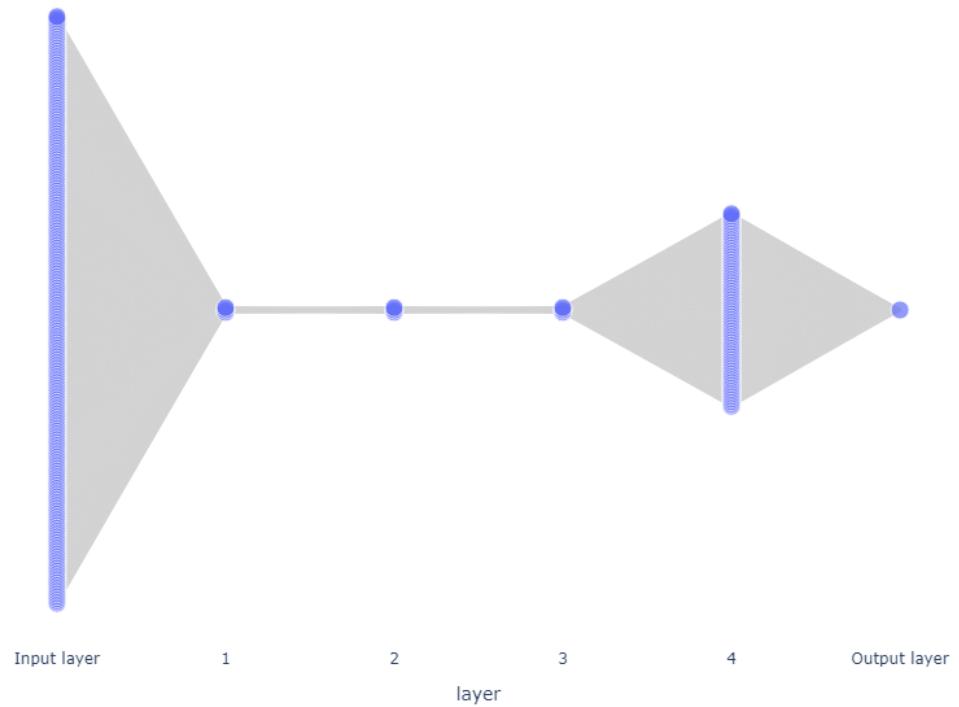


Wind speed (norm) - Label: 0 (train)

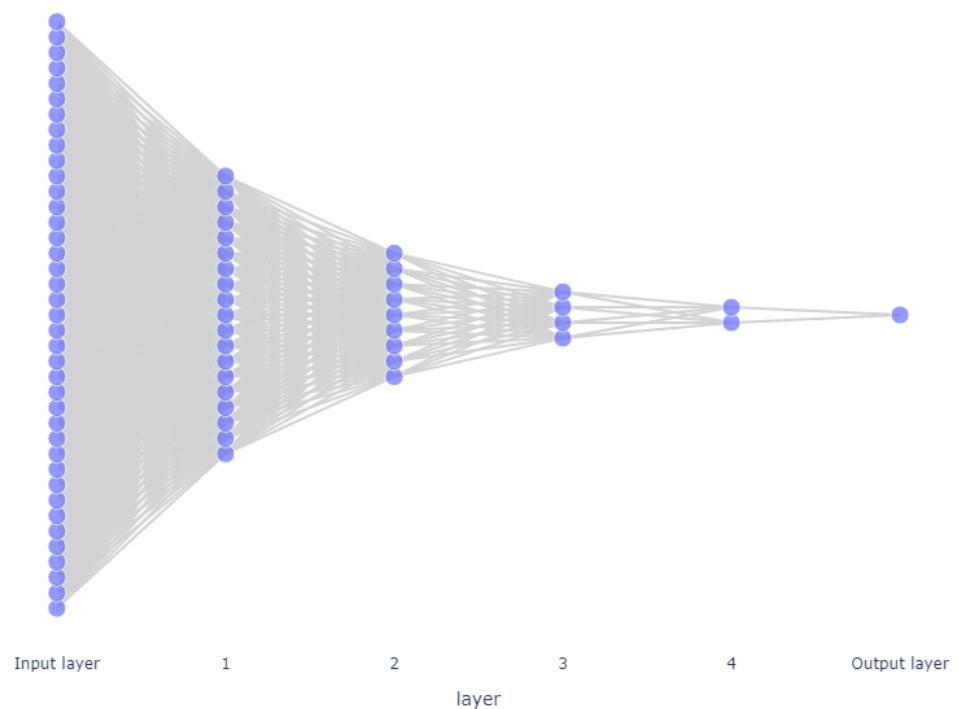


## 11.4 Network Architectures

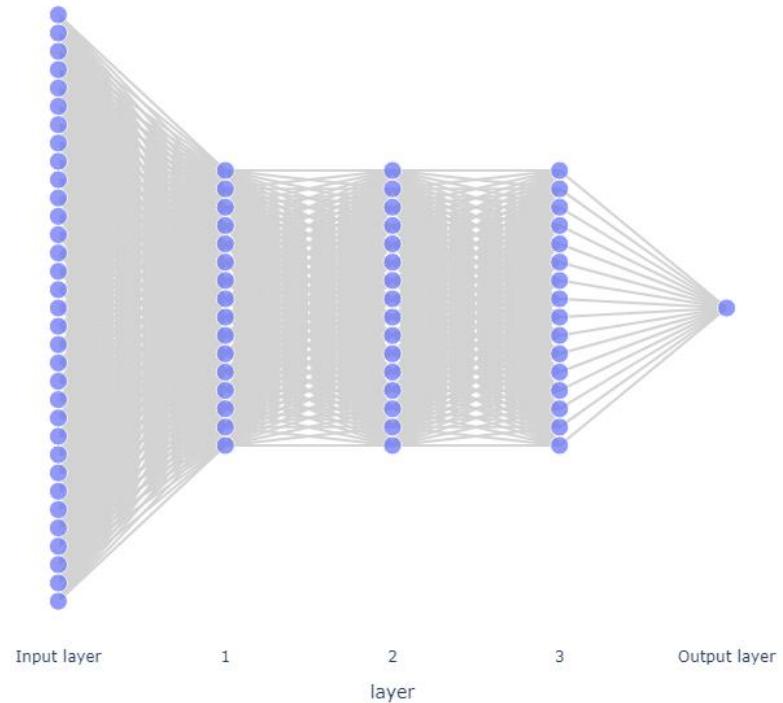
ANN architecture: MLP classifier



ANN architecture: MLP classifier



ANN architecture: RNN classifier



## 11.5 Robust Peak Detection (z-score)

Code source: <https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data>

Python implementation. Required libraries:

- numpy as np

```
def thresholding_algo(y, lag, threshold, influence):
    """Robust peak detection algorithm (using z-scores)

    Args:
        y (_type_): y_vector / time series
        lag (_type_): the lag of the moving window
        threshold (_type_): the z-score at which the algorithm signals
        influence (_type_): the influence (between 0 and 1) of new signals on the mean and

    Returns:
        _type_: dict {
            signals
            avgFilter
            stdFilter
        }
    """

    signals = np.zeros(len(y))
    filteredY = np.array(y)
    avgFilter = [0]*len(y)
    stdFilter = [0]*len(y)

    avgFilter[lag - 1] = np.mean(y[0:lag])
    stdFilter[lag - 1] = np.std(y[0:lag])

    for i in range(lag, len(y)):
        if abs(y[i] - avgFilter[i-1]) > threshold * stdFilter[i-1]:
            if y[i] > avgFilter[i-1]:
                signals[i] = 1
            else:
                signals[i] = -1

            filteredY[i] = influence * y[i] + (1 - influence) * filteredY[i-1]
            avgFilter[i] = np.mean(filteredY[(i-lag+1):i+1])
            stdFilter[i] = np.std(filteredY[(i-lag+1):i+1])
        else:
            signals[i] = 0
            filteredY[i] = y[i]
            avgFilter[i] = np.mean(filteredY[(i-lag+1):i+1])
            stdFilter[i] = np.std(filteredY[(i-lag+1):i+1])

    return dict(signals = np.asarray(signals),
                avgFilter = np.asarray(avgFilter),
                stdFilter = np.asarray(stdFilter))
```

## 12 Declaration of Authorship

I hereby certify that I composed this work as well as the accompanying code completely unaided, and without the use of any other sources or resources other than those specified in the bibliography. All text sections not of my authorship are cited as quotations and accompanied by an exact reference to their origin.

Place, date: Bern, 28.06.2023

Signature:

Joël Tauss