UPPSALA
UNIVERSITET

PROJECT REPORT

Evaluating Locally Measured Weather and Weather Services

Karolina Hedberg, Rebecka Homman & Johan Rideg

Project in Computational Science: Report

January 2021

Department of Information Technology

Abstract

This project was conducted in collaboration with the Gigacow project at the Swedish University of Agricultural Sciences (SLU). The aim of Gigacow is to combine data from multiple sources in order to optimize milk production and welfare of dairy cows. One kind of data that Gigacow are interested in investigating is weather data. For this, they need reliable weather data from the farms included in their network. The aim of this project was to compare interpolated and observed local weather data to conclude whether the interpolated data is a feasible approximation of the observed data. For this, we have investigated the difference between interpolated and observed data of the air temperature, relative humidity, precipitation and wind speed at the locations of some local weather stations. The results showed that the absolute error between observed and interpolated data of the temperature and relative humidity was smaller during fall and winter compared to spring and summer, probably due to more stable weather conditions. For temperature, relative humidity and wind speed we discovered patterns in the diurnal variation. These were suspected to depend on the course of the sun. The results also showed that the interpolated data corresponded best to the observed data for temperature and relative humidity, probably because these parameters are likely to have only small local variations. These might thus be suitable to approximate with the interpolated data. The precipitation and wind speed had larger absolute errors, which was suspected to be due to more complex local behaviours.

Acknowledgements

The authors of this paper would like to extend special thanks to our supervisor Tomas Klingström for guiding us in our decision making and for his unending commitment to this project. We would also extend special thanks to Anne Fouilloux who guided us through Galaxy and provided us with knowledge without this project would not have been possible. Lastly, we would like to thank our contacts at SMHI and LantMet for providing many important insights into their respective data.

Contents

1 Introduction

Over the years, several animal-environment studies have been made connecting milk production and physical welfare of dairy cows with local weather conditions. An example of such a connection is heat stress the effect of which on dairy cows is most commonly measured by a temperature humidity index (THI) [1]. Heat stress is known to cause losses in milk production [2]. It can also cause reductions in fertility and increased embryonic mortality [3]. Other parameters connected to heat stress such as wind speed, rainfall and thermal radiation are often not considered since data of these is usually not easily available [2]. However, a study made on the responses to short-term exposure of rain and wind concluded that these two factors are likely to be important for animal welfare. When exposed to only rain or a combination of rain and wind a decline in lying time, feed intake and skin temperature was observed [4]. Environmental factors connected to local weather conditions have also been connected to the occurrence of mastitis, an inflammation of the udder. Mastitis is the disease that cause most loss within the Swedish dairy production every year [5]. A study from 1988 implies a connection between the occurrence of mastitis and a monthly high THI [6]. A high THI could thus be suspected to have a considerable impact on the dairy production.

Gigacow is an infrastructure investment by the Swedish University of Agriculture Sciences (SLU) to increase the exchange between researchers and industry [7]. Their aim is to combine genetic data of dairy cows with data from the farms' monitoring systems in order to optimize milk production. To achieve this, Gigacow is working on creating a platform for data collection that can be connected to the systems at the dairy farms. Today, 17 Swedish dairy farms are connected to the Gigacow network which results in over 5000 dairy cows forming the basis of the research [7]. This project is a part of Gigacow's new research concerning how animal health and milk production is affected under different weather conditions.

To conduct studies on how milk production and animal health is impacted by different weather conditions, Gigacow require access to reliable weather data. For this they can use either interpolated weather data from the Swedish Meteorological and Hydrological Institute (SMHI) or observed weather data collected by local weather stations located at the farms in the Gigacow network. The purpose of this project was to evaluate the need to install local weather stations at the farms in the Gigacow network by comparing interpolated weather data from SMHI with observed data collected from local weather stations. These locally observed weather data was provided by SLU LantMet. The goal was to investigate the feasibility of using interpolated data to approximate the weather conditions at the Gigacow farms, instead of installing a system of local observation stations. To reach this goal we have investigated the behaviours of four weather parameters including air temperature, relative humidity, precipitation and wind speed.

2 Background

2.1 MESAN (AROME) Dataset

MESAN is a meteorological analysis model that is used by SMHI to describe the current weather situation by a geographic interpolation. The model uses forecast fields as an initial guess and then modifies it with observations from weather stations, satellites and radars to calculate the current weather situation in a grid covering Scandinavia [8, 9]. The MESAN (AROME) dataset consists of archived hourly MESAN data from June 2016 up to the day before the data is accessed [9]. The data is available through a Web API from SMHI and is represented in a GRIB (GRIdded Binary) format with a rotated longitude and latitude grid. For a rotated grid the south pole is relocated from latitude -90 to longitude 0 in order to reduce the differences in distances between the northern and southern parts of the model area [10]. GRIB is a binary gridded data format defined by the World Meteorological Organization (WMO) and is commonly used for the representation of weather data. The grid for this dataset consists of squares with a size of 2.5×2.5km. Each grid point can be thought of as a virtual observation station. A grid point is selected by picking a coordinate and choosing the nearest point on the grid [11]. To handle GRIB data efficiently, you need both a program to open and decode GRIB files and the ability to translate the rotated grid coordinates [10]. MESAN data from the last 24 hours can also be fetched through a Web API from SMHI and is returned in JSON format [11].

2.2 LantMet Dataset

The LantMet database is a collaboration between SLU FältForsk, the Swedish Board of Agriculture and the Rural Economy and Agricultural Societies that stores weather data both from SMHI and local weather stations. The data is mainly used for prognosis models in plant protection [12]. For the local weather stations, there exists a document describing where the station should be placed and how they should be attended to. As an example, the stations should be placed no closer than 30 m to a paved road and should be surrounded by at least 8-10 m^2 short grass. The stations should also be overseed at least once every two weeks and be calibrated regularly [13]. The standard is that the local weather stations observe data every quarter of an hour. From this raw data, hourly data is computed. For missing or obviously erroneous measurements, hourly data is replaced with interpolated data from the SMHI dataset MESAN (AROME) [13]. This replacement is possible for parameters including air temperature, relative humidity, precipitation and wind speed. A significant portion of LantMet stations utilizes a "tipping bucket" rain gauge to record precipitation which can not be reliably operated during freezing temperatures (e-mail LantMet 2020-11-27). If there has been precipitation during a period of freezing temperatures some precipitation is measured during the melting (e-mail LantMet 2020-12-22). Weather data from LantMet can be fetched in JSON format through a Web API. The data includes station ID, station name, coordinates and weather parameters [14].

3 Method

To conduct a comparison between interpolated and observed weather data, we choose to compare data from a set of local weather stations within the LantMet dataset with data from the corresponding locations in the gridded MESAN (AROME) dataset. The local weather stations are chosen in consideration of both their geographic location and the amount of complimentary MESAN (AROME) data in the dataset. All weather stations with a significant amount of MESAN (AROME) data are disregarded. The reasoning is to choose weather stations where at least approximately 90% of the dataset consists of local weather data. However, since we also consider the geographical spread of the stations we choose to use some weather stations where the content of local weather data is slightly lower. In the end we considered weather data from 18 weather stations, the locations of these are visualized in Figure 1. A table of names and station ID:s of the chosen local weather stations from the LantMet dataset is presented in Appendix A. The observed weather data from the LantMet dataset was fetched through their Web API. The corresponding MESAN data was obtained from SMHI by supplying them with a list of coordinates of the local weather stations from the LantMet dataset. SMHI then extracted archived MESAN data from the nearest grid point in the MESAN (AROME) dataset and delivered the data as CSV files.

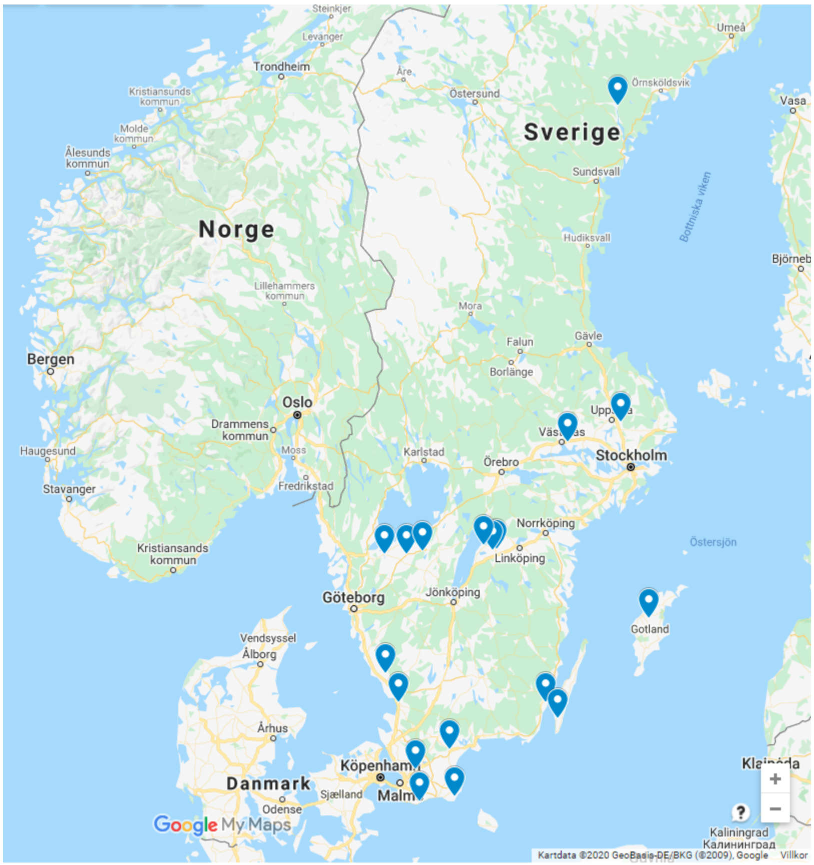

Figure 1: Map of the local weather stations in the LantMet dataset chosen for analysis.

For the data analysis, we investigate the correlation and the absolute difference between the observed and the corresponding interpolated weather data every hour for the coordinate of each chosen station. The years over which the data is compared are chosen with respect to the amount of available observed weather data. In the end we retrieved data from every hour between the beginning of March 2017 and the end of February 2020. To get a perception of the variation in the comparison over the year, the weather parameters are examined separatly for each season. The seasons are divided by using the calender definition from SMHI. The spring is thus considered as March, April and May, the summer as June, July and August, the fall as September, October and November and the winter as December, January and February.

The analysis of the weather data is done in Galaxy / Climate[1] which is a workbench based on the Galaxy framework [15]. Galaxy is an open source software framework aiming to "develop and maintain a system that enables researchers without informatics expertise to perform computation analyses through the web" [16]. Galaxy / Climate provides tools for analysis of weather data. We use the tool "Interactive Climate Notebook" which is based on Jupyter and contains pre-installed

[1]The Galaxy server that was used for some calculations is in part funded by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)).

tools for climate science [15].

4 Results

Figure 2 presents box plots showing the distribution of the absolute error between the observed and interpolated data for each station, season and investigated parameter. The box plots are visualized without outliers in order to avoid cluttered plots and to get a clearer view of the general behaviour in the data. By observation of the temperature plots it is clear that both the median and distribution of the absolute error are in general reduced during the fall and winter compared to the spring and summer. In the summer, the medians are in general below 2°C while in the winter, most medians are below 1°C. The same behaviour is also true for the relative humidity. For this parameter, the medians of the absolute errors are below 10 percentage points in the spring while in the winter the medians for most stations are below 5 percentage points. For the precipitation plots all absolute errors when both the observed and interpolated data had zero precipitation were removed. Otherwise all medians were computed to be zero since the precipitation was zero in the majority of the data points. For the box plots of the absolute error of precipitation and wind speed, no obvious behaviours concerning the differences between neither stations nor seasons can be seen.

From the box plots in Figure 2 it can also be noticed that the absolute error for some stations deviates from the others, especially concerning the temperature and relative humidity. The distribution of the absolute error in summer and fall temperature for station 23659 is comparatively large. This LantMet weather station collected observed data mainly during the summer months of the investigated time period. For the rest of the year the dataset for this station consists exclusively of MESAN (AROME) data. Thus, the deviation in absolute error for summer and fall may indicate irregularities in the observations from this station. When observing the figures containing the box plots for the temperature in the summer it can be seen that also station 40145 has a larger distribution of the absolute error than the rest of the stations. Similarly, in the figures presenting the box plots for the relative humidity in the spring and summer, it can be noticed that the stations 40013 and 25652 has a comparatively large distribution of the absolute errors. These four mentioned stations were all discarded for the further analysis in order to avoid misleading results.

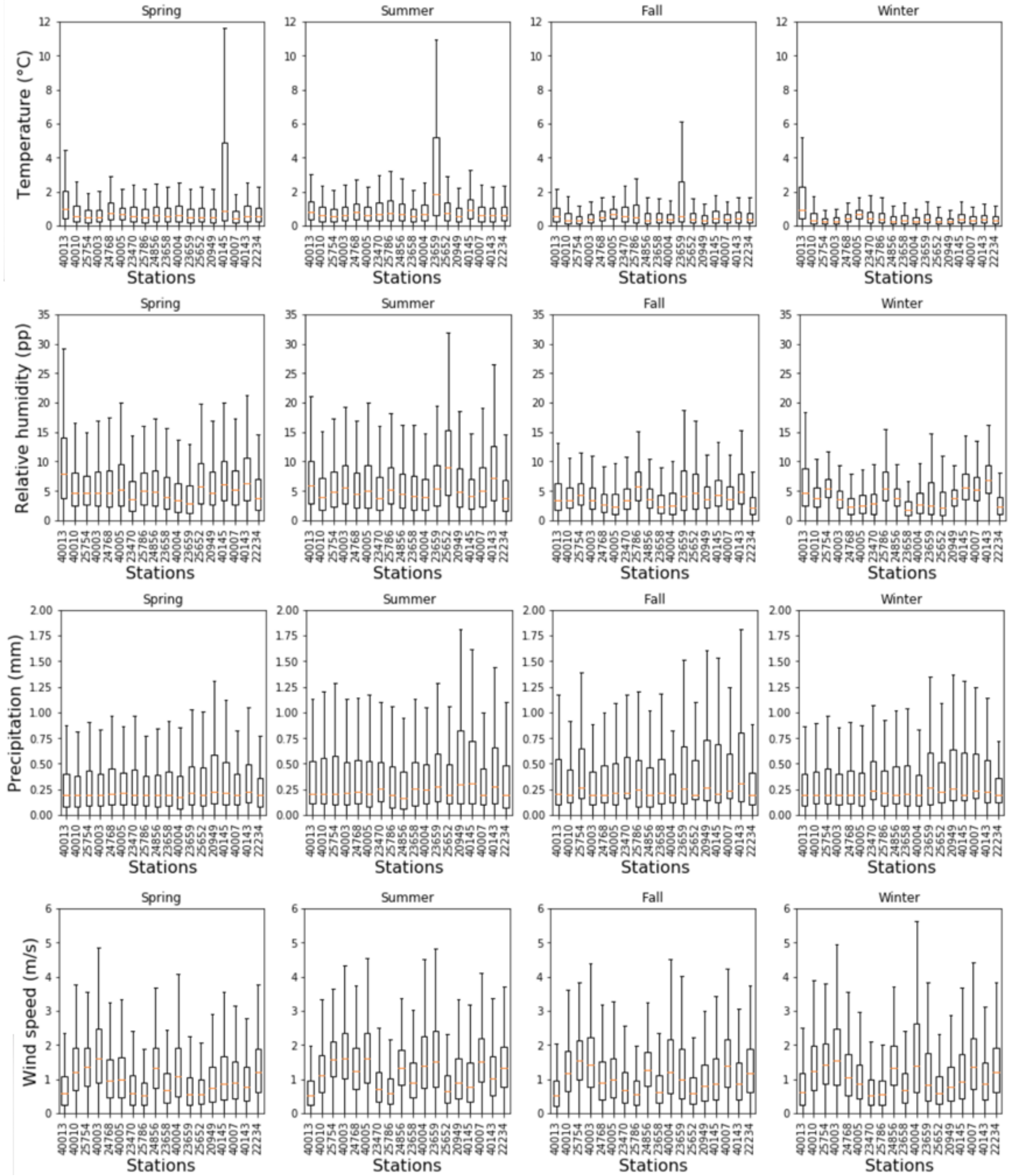

Figure 2: Box plots presenting the distribution of the absolute error between the observed and interpolated data for each station, season and parameter when outliers were disregarded.

In Figure 3 the diurnal variation of the absolute error between the observed and interpolated data for each station, season and parameter is presented. The plots shows how the absolute error of the parameters varies over a average day in every season during the investigated time period. The diurnal variation was computed by calculating the mean value of the absolute error for each hour of the day in each season. When looking at the diurnal variation of the absolute error for the temperature it can be seen that during an average day, it has two peaks. The distance between these peaks is larger in the spring and summer than in the fall and winter. From the plots presenting the absolute error of the relative humidity it can be noticed that it has one peak during the day. The width of the peak is larger in the spring and summer than in the fall and winter. In the figures presenting the precipitation, no clear patterns in the diurnal variation exists. However, it can be noticed that the absolute error of the precipitation is smaller than what was observed in the box plots in Figure 2. This depends on that the absolute errors when both the observed and interpolated data had zero precipitation were included in this figure. Thus the average absolute error was expected to be close to zero. Regarding the diurnal variation of the wind speed, it can be noticed that the absolute error for most of the stations seems to decrease during an average day in spring and summer. During the fall and winter the absolute error seems to decrease for some stations during the day but not as noticeably as during the spring and summer. A general behaviour that appears most clearly in the plot for the relative humidity is the constant offset of the absolute error. In the plot of the absolute error in the winter, it can be seen that the pattern of the error is the same for all stations but that they are displaced in different heights.

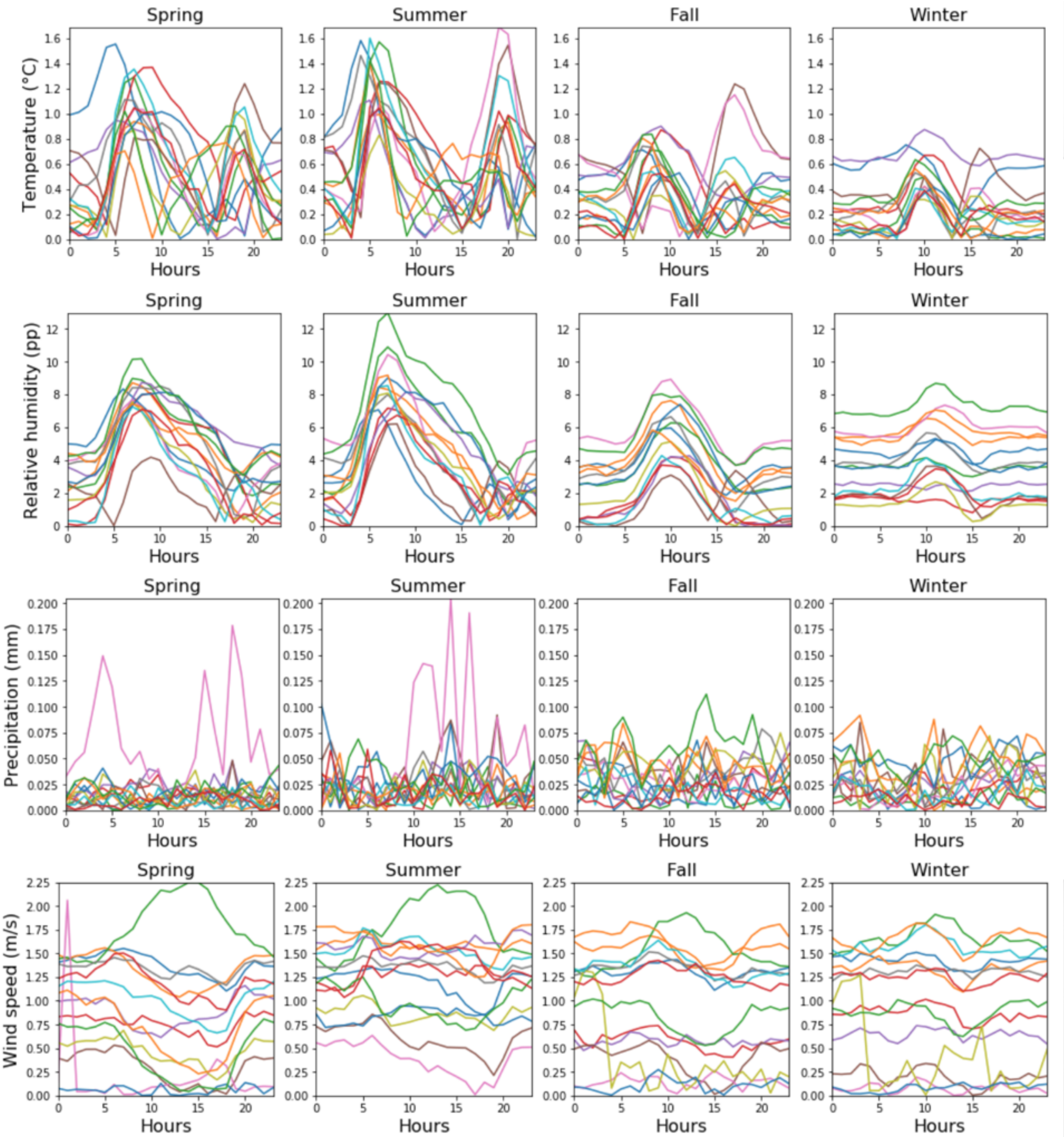

Figure 3: Average absolute error between the observed and interpolated data for each hour of the day for each station, season and parameter.

Figure 4 contains scatter plots of the interpolated data from MESAN (AROME) and observed data from the LantMet dataset for all seasons and parameters for the local LantMet station

40004. This is used as an example station since it has some behaviours that recurs for several of the investigated stations. Similar plots for the rest of the stations are attached in Appendix B. Figure 4 contains both the data points and a linear regression model based on the data. The model is given on the form $y = kx+m$, where x is the interpolated MESAN (AROME) data and y is the observed data from the LantMet dataset. The black dashed line visualizes what the model would look like if the data matched completely. From Figure 4, interpolations and observations are seen to correlate best for temperature and relative humidity since the data points does not vary as much from the linear regression models compared to other parameters. By inspection of the regression model for the temperature, it is also clear that this model corresponds best to the ideal case compared to the other weather parameters. This since the regression models for temperature is closest to the case $y = x$ for all seasons. Second best correlates the interpolations and observations for relative humidity. Compared to the temperature, the data points of the humidity vary slightly more from the corresponding linear regression models. The humidity is consistently reported higher in observations than interpolations across all stations and seasons as the linear model consistently have coefficients $[k, m] = [\approx 1, > 0]$. These observations, concerning the correlation between observed and interpolated temperature and relative humidity, is similar between most of the investigated weather stations, as can be seen from the scatter plots in Appendix B. Observations and interpolations of the precipitation correlate relatively poorly, something that is also seen to be true for all stations. It is however seen that the precipitation is prone to having high observed values in regards to low interpolated values. In Figure 4, the wind speed seem to correlate in "tails", having a negative impact in the overall correlation. The tails are consistently indicating lower observations relative to interpolations. Otherwise for stations not showing any tails, the correlation of wind speed is significantly better as can be seen from the figures in Appendix B.

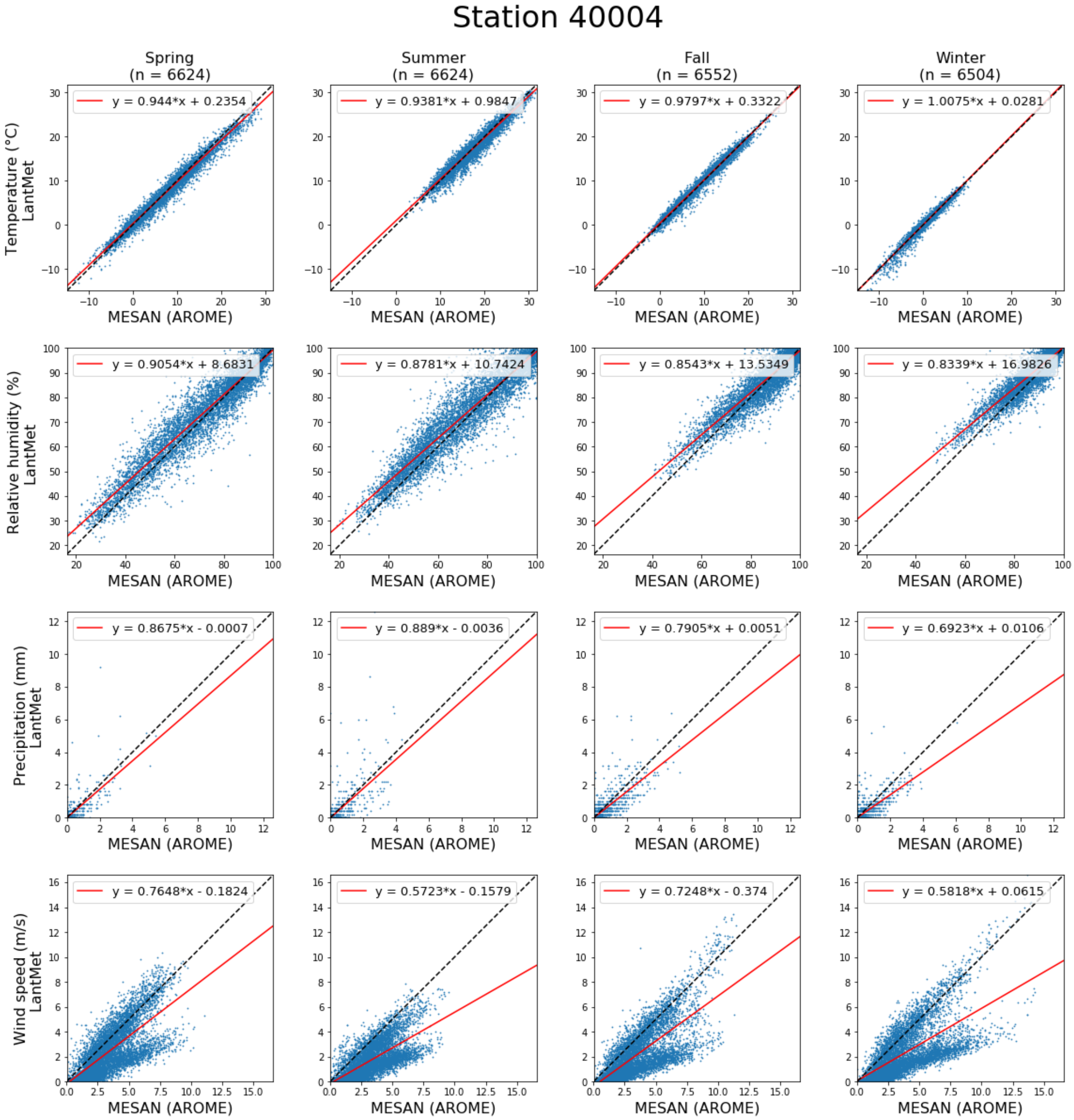

Figure 4: Scatter plots of interpolated data from MESAN (AROME) and observed data from LantMet for all four seasons and parameters. The red line shows a linear regression model based on the data while the black dashed line represents the ideal case.

5 Discussion

5.1 Discussion of results

From the results presented in Figure 2 and Figure 3 it is seen that the absolute error for both temperature and humidity generally produces lower and less varied errors during the fall and winter compared to the spring and summer. One possible explanation for this, is that the weather may be more stable during these months whereas the spring and summer may exhibit a more fluctuating weather. This would cause the interpolated data to be a better approximation of the local weather during fall and winter. The diurnal variation of the absolute temperature error also seems to follow some general pattern. As was noted in the results, the absolute error has two peaks. Since the peaks are further apart during spring and summer and approach each other at fall and winter it is reasonable to argue that the peaks somehow correspond to sunrise and sunset. A similar behaviour can be observed for the relative humidity. However, the absolute error of this parameter only has one peak and thus rather seems to correspond to the hours when the sun is up. This is also implied by the observation that the width of the peaks shrinks during the fall and winter when the days are getting shorter. By observation of the plots of the diurnal absolute error of the wind speed, a similar behaviour can be distinguished mainly during the spring and summer. Although, in this case the peak is inverted such that the absolute error decreases. The absolute error in wind speed is thus also suspected to depend on the sun. Since this behaviour of the absolute error can be seen throughout all stations it is reasonable to argue that this may depend on the interpolated data rather than on the locally observed data.

From the diurnal plots in Figure 3 it is also noted that the absolute error of the relative humidity has different offsets in absolute error for different stations. This indicates that there exists a systematic error between the observed and interpolated data. This may both depend on the fact that the interpolated data is fetched from a grid point that does not correspond directly to the location of the local weather station and on the calibration and location of the local weather station.

For the average diurnal behaviour, the absolute error of the precipitation falsely indicates a good interpolation, as seen in Figure 3. This is because more than 85 % of data points have zero local and interpolated precipitation simultaneously, resulting in a low mean error. Precipitation is somewhat of a special parameter in the sense that it is virtually always zero in both data sets. Therefore, it may be more fair to investigate the absolute error when removing the data points where both observed and interpolated data indicate a zero precipitation as done in Figure 2.

From the scatter plots, it is concluded that observations and interpolations of the temperature correlate relatively well in regard to other parameters across all stations. This is possibly because it may be more easily modelled and not as sensitive to local factors as wind speed and precipitation are suspected to be. Observations and interpolations of the relative humidity correlate somewhat well. However, the observations tend to be higher across all stations and seasons. The reasons behind this may be local factors or hidden within the interpolation model as this systematic error is apparent across all stations. Precipitation is prone to having high observed values in regards to low interpolated values. This is most likely due to local rain showers that are not captured in the interpolated MESAN (AROME) data. As stated in the results concerning the correlation of the absolute error in wind speed, some stations show several tails where observed tends to be lower than interpolated wind speed. This may be explained by local obstructions, lowering the wind speed in certain directions. Further investigations of how data from individual weather stations is affected by its environment can be used to motivate this theory but are beyond the scope of this study.

5.2 Discussion of choice of methods

A large part of the project consists of investigating methods for collecting interpolated weather data from SMHI. One way is to fetch JSON data directly from one of their open data API:s. This data can be trivially read and manipulated but the drawback is that it only stretches 24 hours back in time. Another alternative is to fetch archived grid data in the form of GRIB files. The problem with this data is that, being in a binary format does not allow for trivial data access and manipulation. The initial strategy was to sample data from the 24 hour MESAN API two times per day and collect a data set ourselves which was done successfully. The problem with this method is that it would result in a small dataset containing weather data from only a few weeks. However, this served as a backup data set if no larger data set could be obtained. Through Galaxy, we learned to extract parameter data from GRIB files. However, this process is slow primarily as a fetch need to happen for each hour to be included in a data set. The GRIB files are relatively large and one file containing 24 hours of data take up about 3.1 GB of storage, further adding to the load of the host server. Regardless, a trial data set was created to serve as a foundation for data analysis and visualisation. Running the program responsible of fetching and extracting data from SMHI GRIB archive for one coordinate take close to two days. Therefore, we would have to restrict ourselves to a smaller set of coordinates in our analysis. In the end data from the MESAN (AROME) dataset was ordered from SMHI and delivered in CSV format which enabled the process of analysing the data through a Jupyter notebook without handling the GRIB format.

The procedure of selecting weather stations from the LantMet dataset can be regarded as somewhat arbitrary. The reason for this is that there is no straight forward way of checking whether observed data has been complemented with MESAN (AROME) data. This was done manually by browsing through all data and estimating the amount of complemented data in the dataset. This method leaves room for subjectivity since it is difficult to approximate exactly how much of the total data that is complemented. An alternative method would be to use the raw data where the dataset has not been complemented with any MESAN (AROME) data. The drawback with this would be that the data might contain unreasonable values since it has not been checked for erroneous measurements. It is also possible that some time points would miss data completely. This would give us a messier data set and would demand more time for cleaning the data, something that could not be done during the allotted time for this project. As described in the method, the seasons are defined by the calendar definition of SMHI. An alternative way would be to use the meteorological definition were the beginning of a season depends on the daily temperatures. This might give a more accurate analysis of the difference between different seasons since the interval of the seasons varies throughout the country.

6 Conclusions and Further Work

From this project it is concluded that the absolute error between interpolated and observed temperatures and relative humidities seem to be smaller during fall and winter compared to spring and summer, something that is suspected to be due to a more stable weather during the fall and winter months. For the diurnal variation, the absolute error of both temperature, relative humidity and wind speed are suspected to be dependent on the sun. The absolute error of the temperature seem to be connected to sunset and sunrise while the absolute error of relative humidity and wind speed rather seem to be dependent on the hours when the sun is up. From the analysis of the correlation between the interpolated and observed weather data it can be concluded that some parameters can be better interpolated, probably because these are less inclined to have local variations. Temperature and relative humidity are shown to be approximated rather well by interpolated data while precipitation and wind speed prove to have more complex local behaviours, making them difficult to interpolate.

Concerning the need to install local weather stations on the Gigacow farms, this is highly dependent on the requirements of the preciseness in each parameter. Temperature and relative humidity prove to be the easiest to interpolate and might thus be suitable to approximate by interpolated data. By the results of this project, precipitation and wind speed prove to be more difficult to interpolate. However, to conclude whether to use interpolated data it should be investigated how sensitive dairy cows are to the effects of small weather variations and thus how large measurement errors that are acceptable. Concerning the speculations on systematic errors in the MESAN (AROME) dataset, more information can hopefully be obtained by contacting a meteorologist with knowledge of the MESAN model. Similarly, responsible at LantMet can be contacted for information about the measurement errors in the LantMet dataset.

A suggestion for a continued study is to compare the interpolated and observed values of the wind speed in different directions. Thus enabling an analysis of how the observed wind speed can be affected by local obstructions and how this affect the feasibility to approximate the wind speed with interpolated data. Another result that may be interesting to investigate further is how the diurnal variation of the absolute error of temperature, relative humidity and wind speed depend on the sun. A method to conduct an examination of this would be to offset the diurnal plots in Figure 3 such that the hour axis begins at sunrise. This would give a clearer indication on whether there is a connection between the absolute errors and sunrise.

References

[1] Dikmen S, Hansen P.J. Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment? Journal of Dairy Science. 2009;92(1):109-16. Available from: https://www.sciencedirect.com/science/article/pii/S0022030209703157

[2] Bohmanova J, Misztal I, Cole J.B. Temperature-Humidity Indices as Indicators of Milk Production Losses due to Heat Stress. Journal of Dairy Science. 2007;90(4):1947-56. Available from: https://www.sciencedirect.com/science/article/pii/S0022030207716818

[3] Hansen P.J. Exploitation of genetic and physiological determinants of embryonic resistance to elevated temperature to improve embryonic survival in dairy cattle during heat stress. Theriogenology. 2007;68(1):S242-9. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0093691X07001367

[4] Schütz KE, Clark KV, Cox NR, Matthews LR, Tucker CB. Responses to short-term exposures to simulated rain and wind by dairy cattle: time budgets, shelter use, body temperature and feed intake. Animal Welfare. 2010;19(4):375-83. Available from: https://www.researchgate.net/profile/Lindsay_Matthews/publication/263077739_Responses_to_short-term_exposure_to_simulated_rain_and_wind_by_dairy_cattle_Time_budgets_shelter_use_body_temperature_and_feed_intake/links/54cf5e8f0cf24601c0933f8b/Responses-to-short-term-exposure-to-simulated-rain-and-wind-by-dairy-cattle-Time-budgets-shelter-use-body-temperature-and-feed-intake.pdf

[5] Lund E, Swedish Board of Agriculture. Djurhälsa 2019 [Internet]. No place of publication: Jordbruksverket; 2020. Djurhälsa. [cited 2020-12-29]. Available from: https://jordbruksverket.se/om-jordbruksverket/jordbruksverkets-officiella-statistik/jordbruksverkets-statistikrapporter/statistik/2020-10-20-djurhalsa-2019

[6] Morse D, DeLorenzo M.A, Wilcox C.J, Collier R.J, Natzke R.P, Bray D.R. Climatic Effects on Occurrence of Clinical Mastitis. Journal of Dairy Science. 1988;71(3):848-53. Available from: https://www.sciencedirect.com/science/article/pii/S0022030288796265

[7] Swedish University of Agricultural Sciences. Gigacow - en SLU-infrastruktur [Internet]. No place of publication: Swedish University of Agricultural Sciences; date unknown [updated 2020-11-12; cited 2020-12-29]. Available from: https://www.slu.se/institutioner/husdjursgenetik/forskning/gigacow/

[8] Häggmark L, Ivarsson KI, Olofsson PO. MESAN - Mesoskalig analys. Norrköping: Sveriges meteorologiska och hydrologiska institut; 1997. RMK No.75. Available from: https://www.smhi.se/polopoly_fs/1.36794!/RMK_75.pdf

[9] Swedish Meteorological and Hydrological Institute. Meteorologisk analysmodell MESAN (AROME) - historiska analysdata [Internet]. No place of publication: Swedish Meteorological and Hydrological Institute; 2016 [updated 2020-09-25; cited 2020-12-29]. Available from: https://www.smhi.se/data/utforskaren-oppna-data/meteorologisk-analysmodell-mesan-arome-historiska-analysdata

[10] Swedish Meteorological and Hydrological Institute. GRIB-format [Internet]. No place of publication: Swedish Meteorological and Hydrological Institute; 2013 [updated 2018-12-18; cited 2020-12-29]. Available from: https://www.smhi.se/data/oppna-data/information-om-oppna-data/grib-format-1.30761

[11] Swedish Meteorological and Hydrological Institute. Get Point Analys [Internet]. No place of publication: Swedish Meteorological and Hydrological Institute; date unknown [cited 2020-12-29]. Available from: http://opendata.smhi.se/apidocs/metanalys/get_analys.html

[12] Swedish University of Agricultural Sciences. Väderdata [Internet]. No place of publication: Swedish University of Agricultural Sciences; 2020 [cited 2020-12-29]. Available from: https://www.slu.se/fakulteter/nj/om-fakulteten/centrumbildningar-och-storre-forskningsplattformar/faltforsk/vader/

[13] Leuchovius T, Karlsson S, Nissen K. Väderstationer i Lantmet – montering, utplacering och skötsel. No place of publication: SLU Fältforsk; 2016 [cited 2020-12-29]. Available from: https://www.ffe.slu.se/lm/ManualLantmet.pdf

[14] Skog TE. Specification for external weather data source for VIPS. No place of publication: Bioforsk Plant Health; 2015 [cited 2020-12-29]. Available from: https://www.ffe.slu.se/lm/json/JSON-Specifikation.pdf

[15] Galaxy / Climate. Welcome to the climate science community [Internet]. No place of publication: UseGalaxy.eu; date unknown [updated date unknown; cited 2020-12-29]. Available from: https://climate.usegalaxy.eu/

[16] Galaxy Project. Galaxy 101 - What is Galaxy? [Internet]. No place of publication: Galaxy Project; date unknown [updated date unknown; cited 2020-12-29]. Available from: https://galaxyproject.org/tutorials/g101/

Appendix

Appendix A: LantMet weather stations

Table 1 presents the station ID, station name and coordinates for each of the 18 investigated local weather stations from the LantMet dataset.

Station ID	Station name	Coordinates (lat, long)
40013	Undrom-Lännäs	63.14682, 17.75392
40010	Uppsala Funbo-Lövsta	59.83020, 17.80930
25754	Brunnby	59.61325, 16.66468
40003	Skänninge	58.40630, 15.10070
24768	Bjälbo	58.38640, 15.00407
40005	Vadstena	58.44270, 14.80780
23470	Logården	58.33865, 12.63607
25786	Lanna	58.34721, 13.12721
24856	Götala	58.38054, 13.47516
23658	Vassmolösa	56.60055, 16.18111
40004	Hallfreda	57.60030, 18.44170
23659	Smedby (only summer)	56.40236, 16.42808
25652	Sannarp Falkenberg	56.94146, 12.66630
20949	Lilla Böslid	56.59782, 12.95124
40145	Hviderup	55.77791, 13.32169
40007	Sandby gård	55.43950, 14.18270
40143	Lovisero	55.37719, 13.40477
22234	Helgegården	56.01990, 14.06210

Table 1: Station ID, Station name and coordinates for analysed weather stations from the LantMet dataset.

Appendix B: Scatter plots for all weather stations

This Appendix contains scatter plots similar to Figure 4 for all weather stations with ID:s given as in Table 1 in Appendix A.

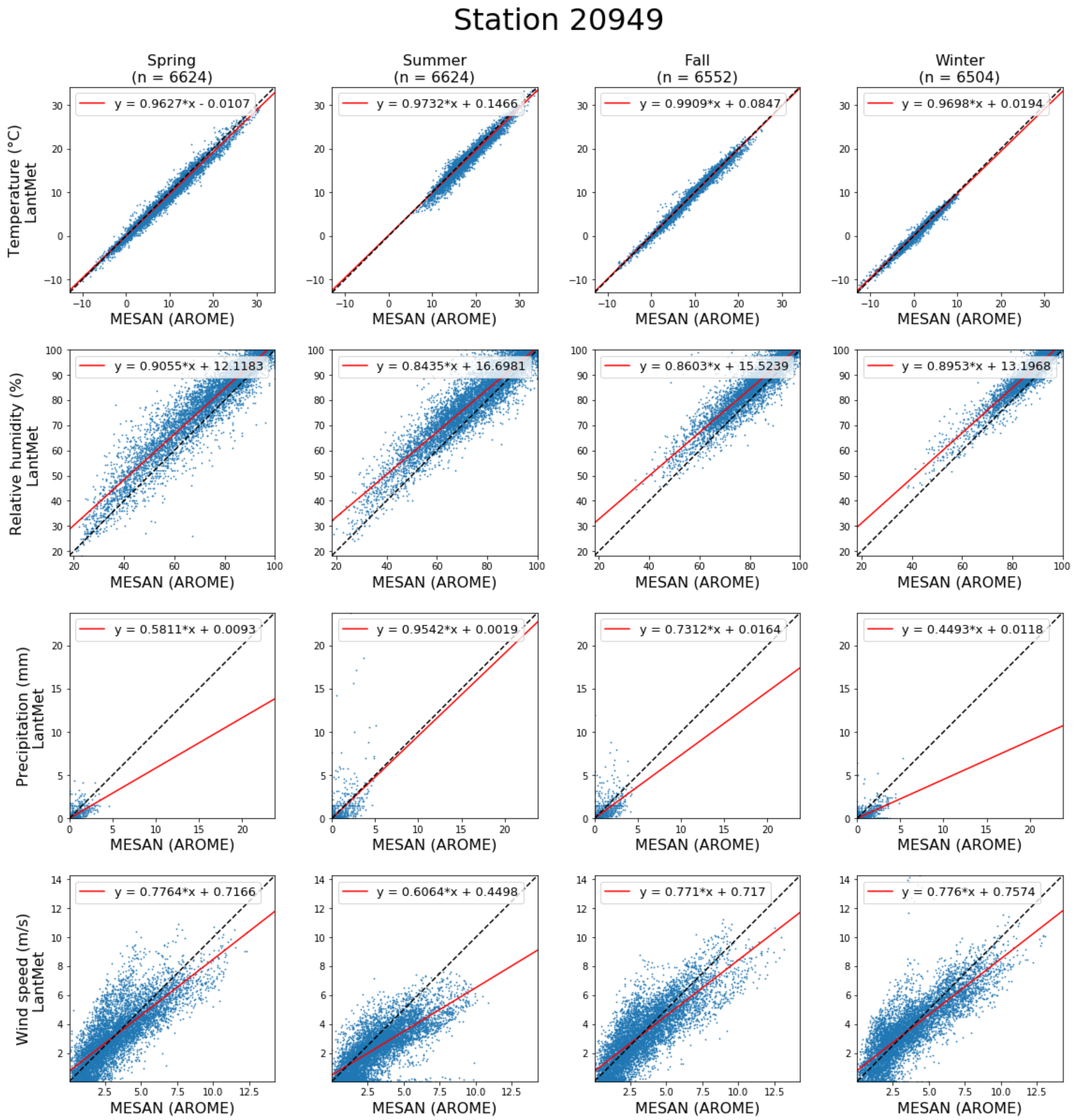

Figure 5: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

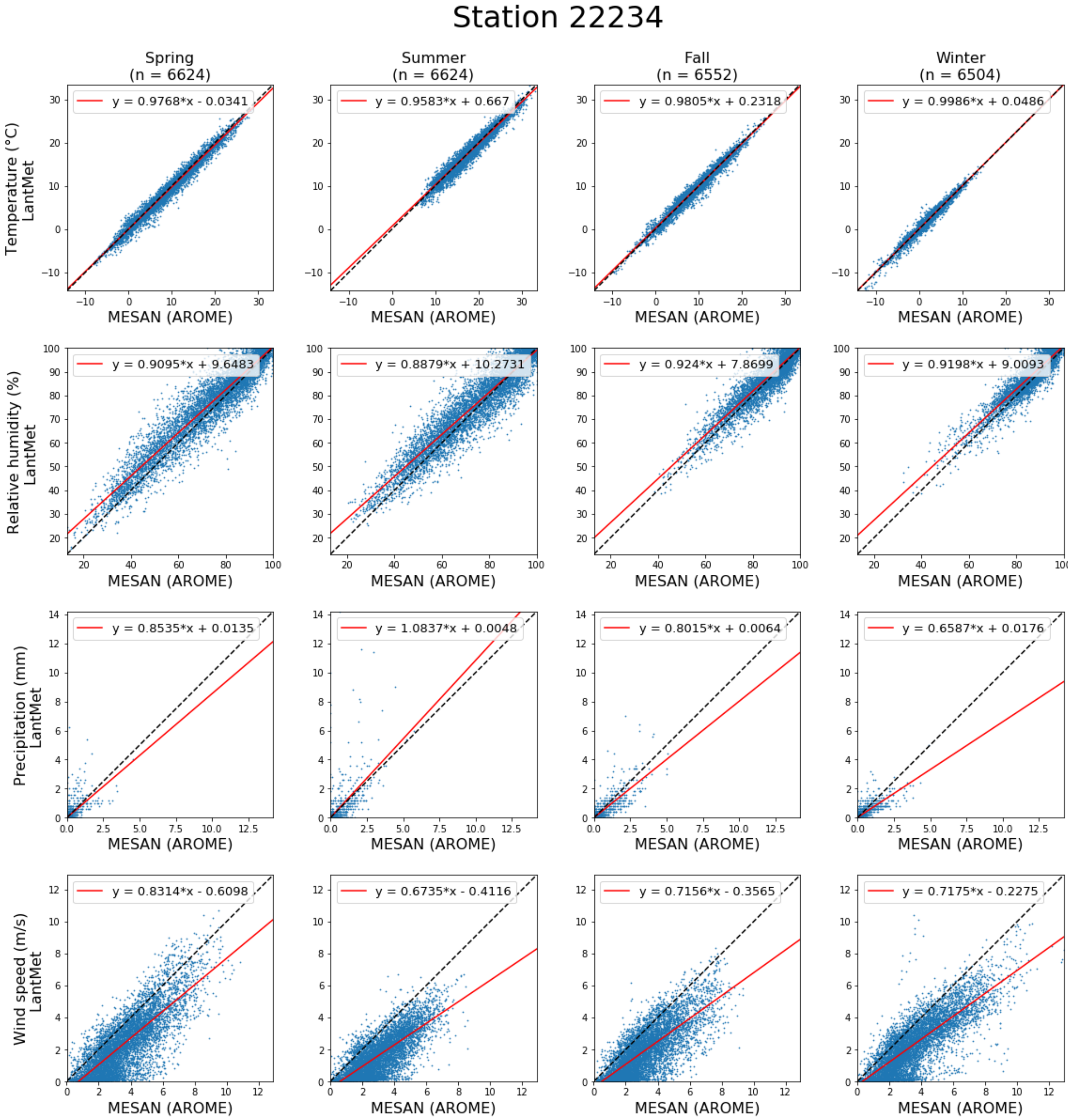

Figure 6: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

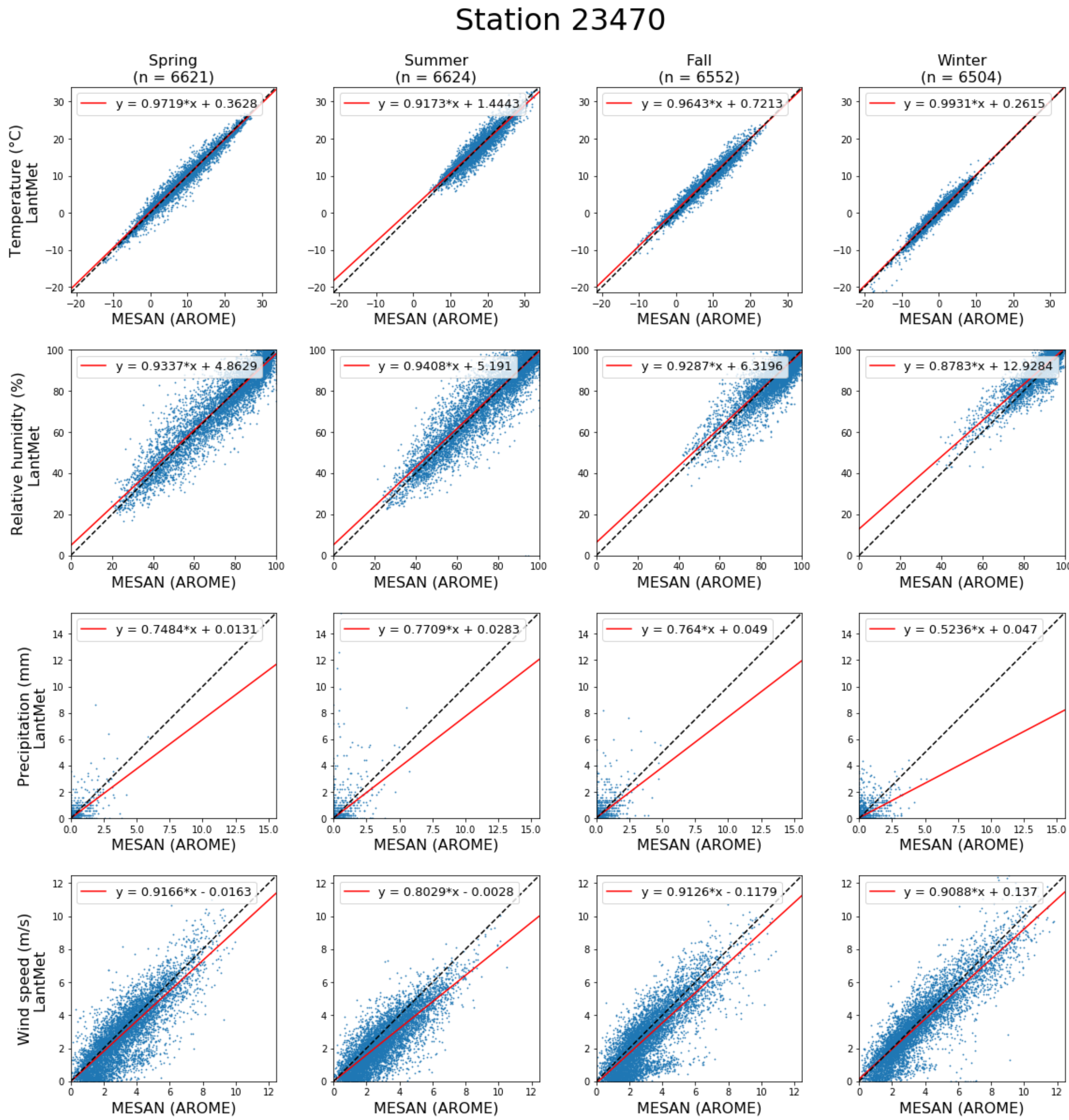

Figure 7: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

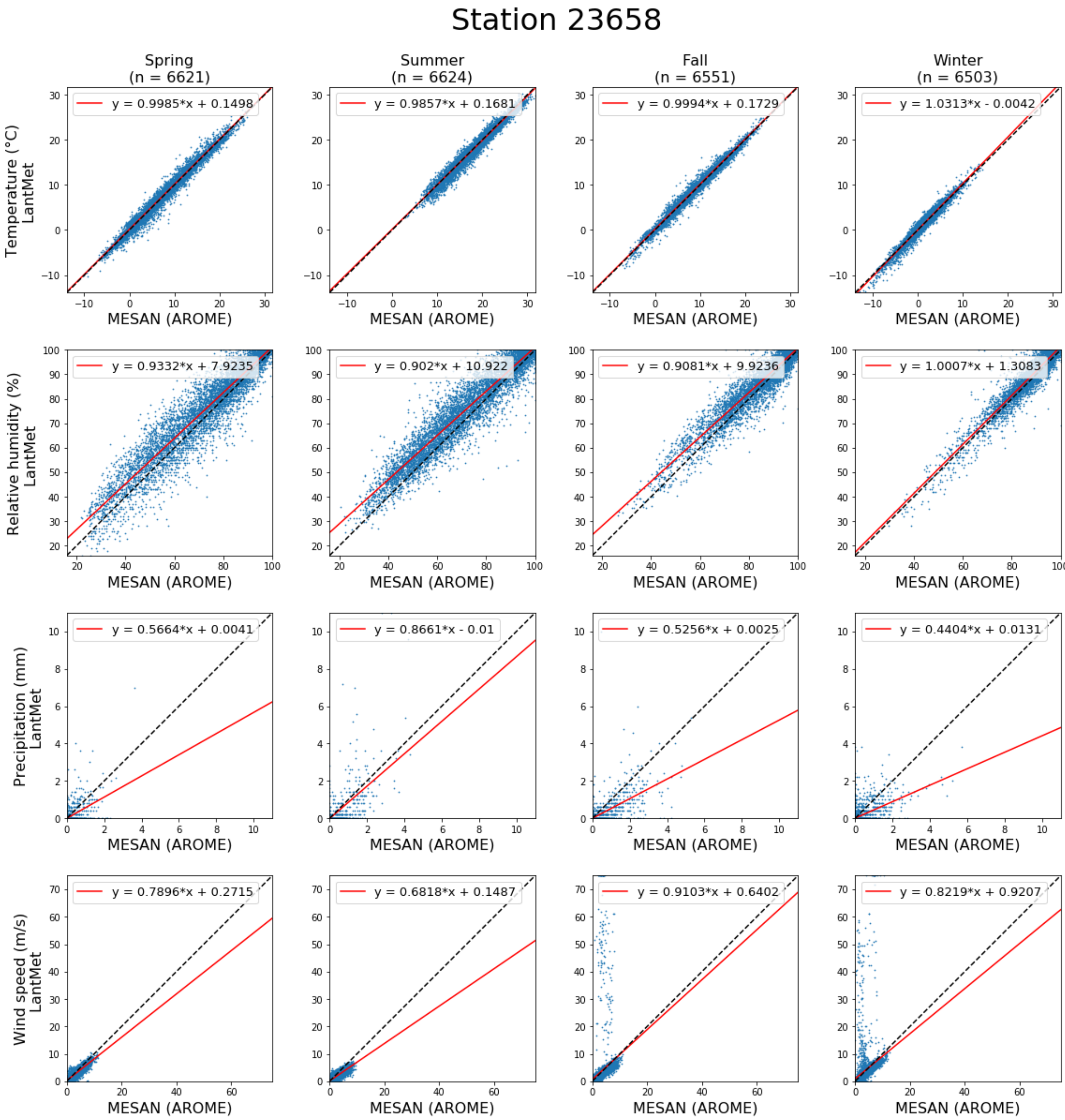

Figure 8: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

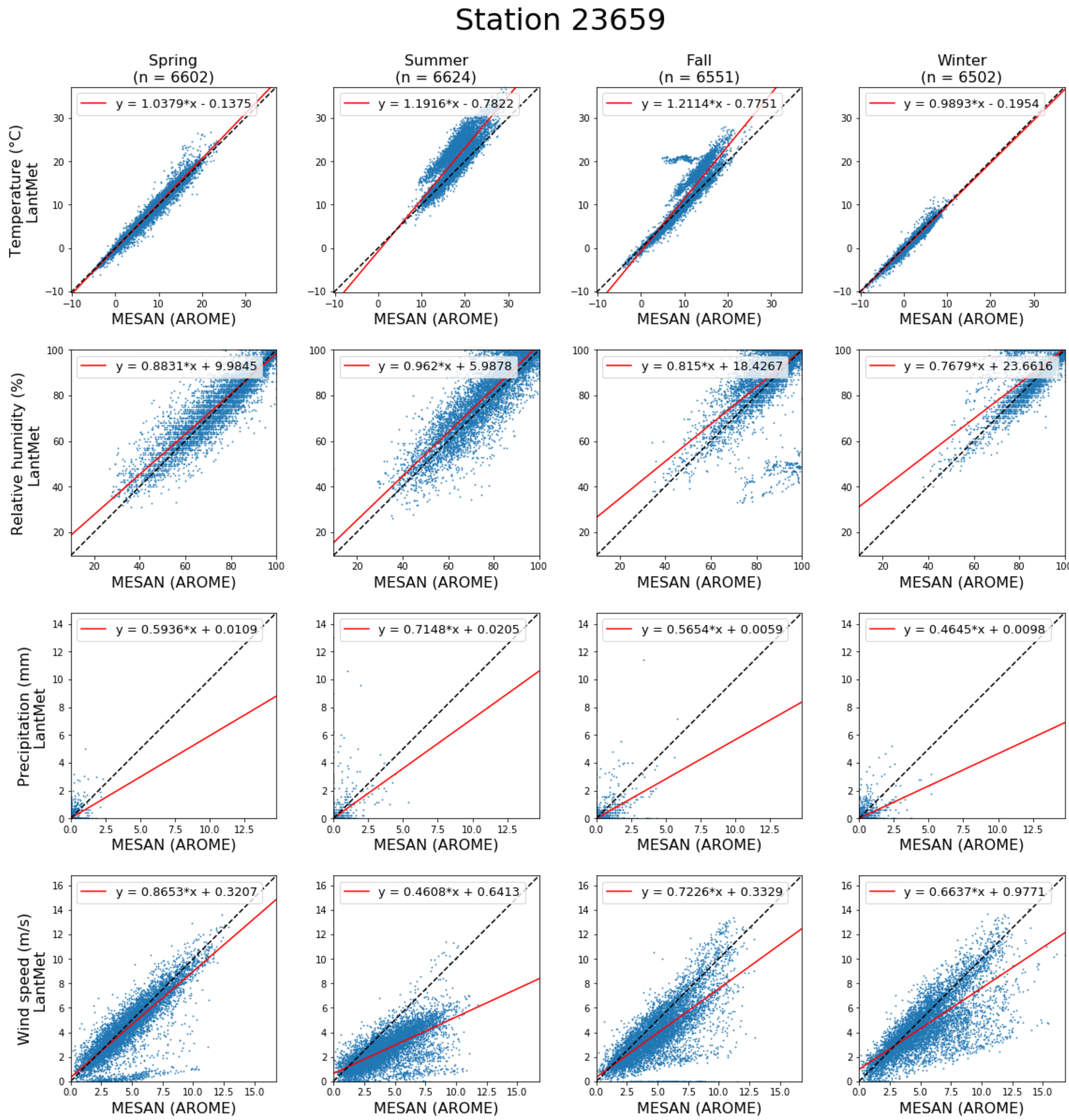

Figure 9: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

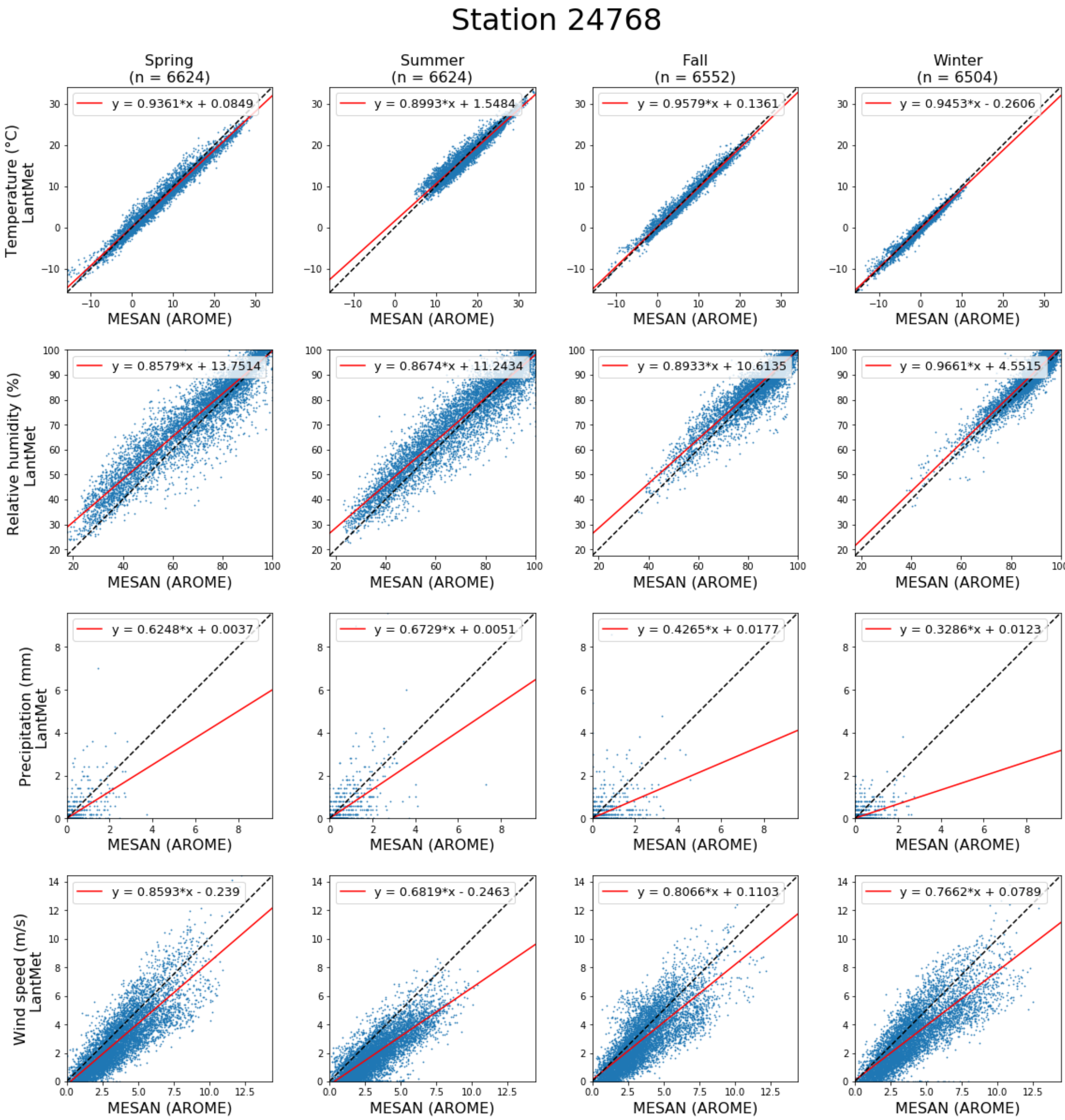

Figure 10: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

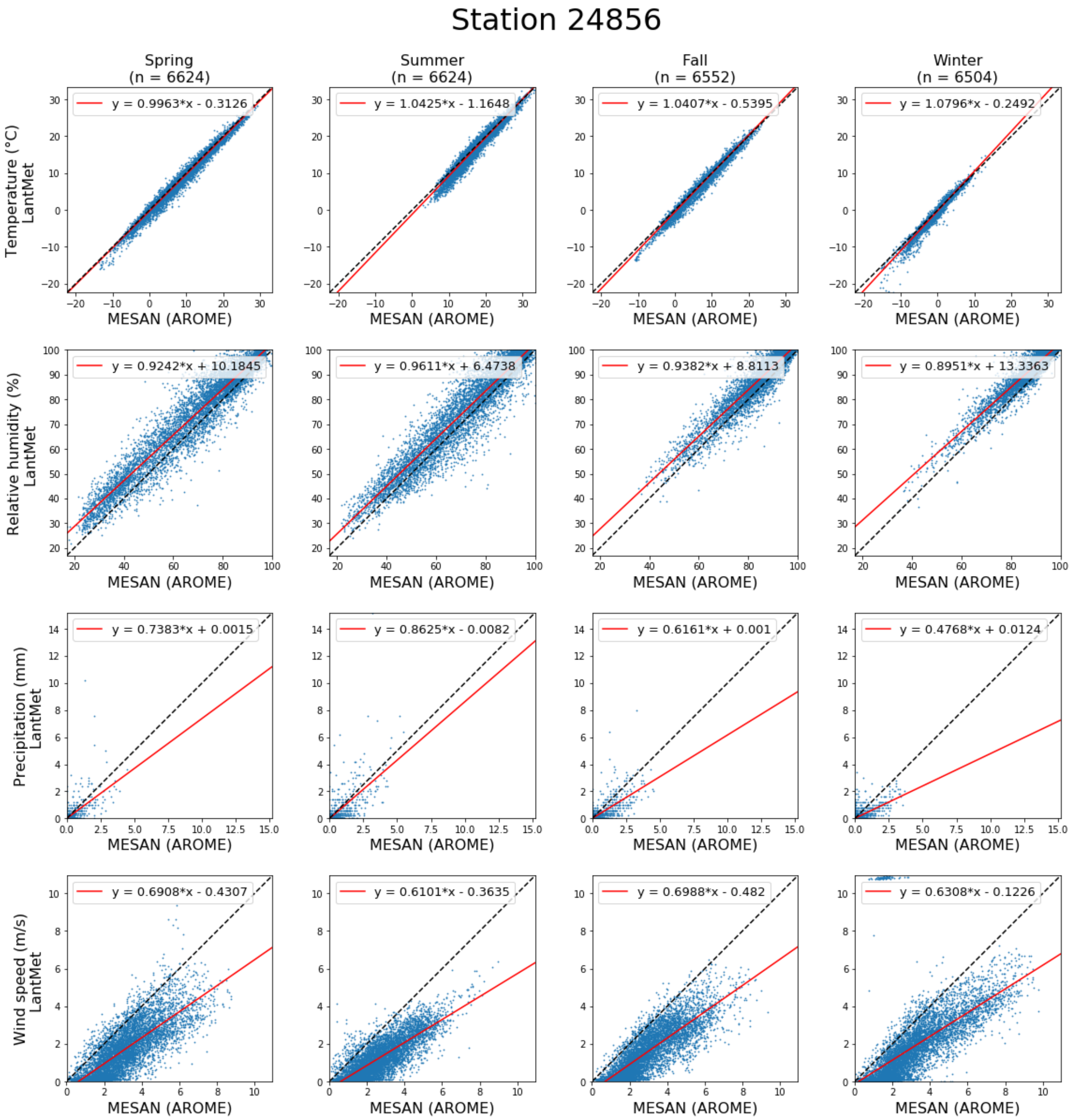

Figure 11: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

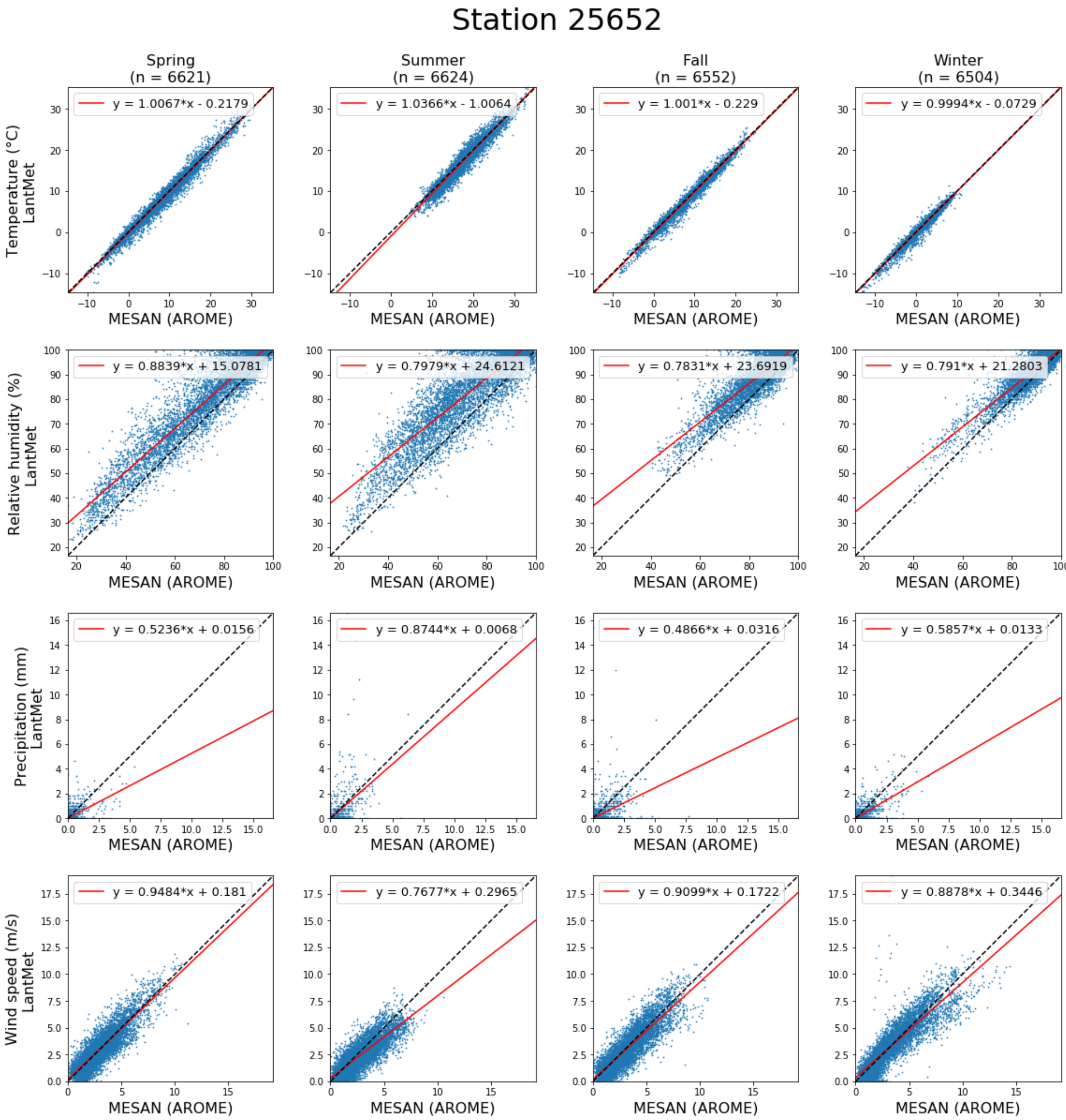

Figure 12: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

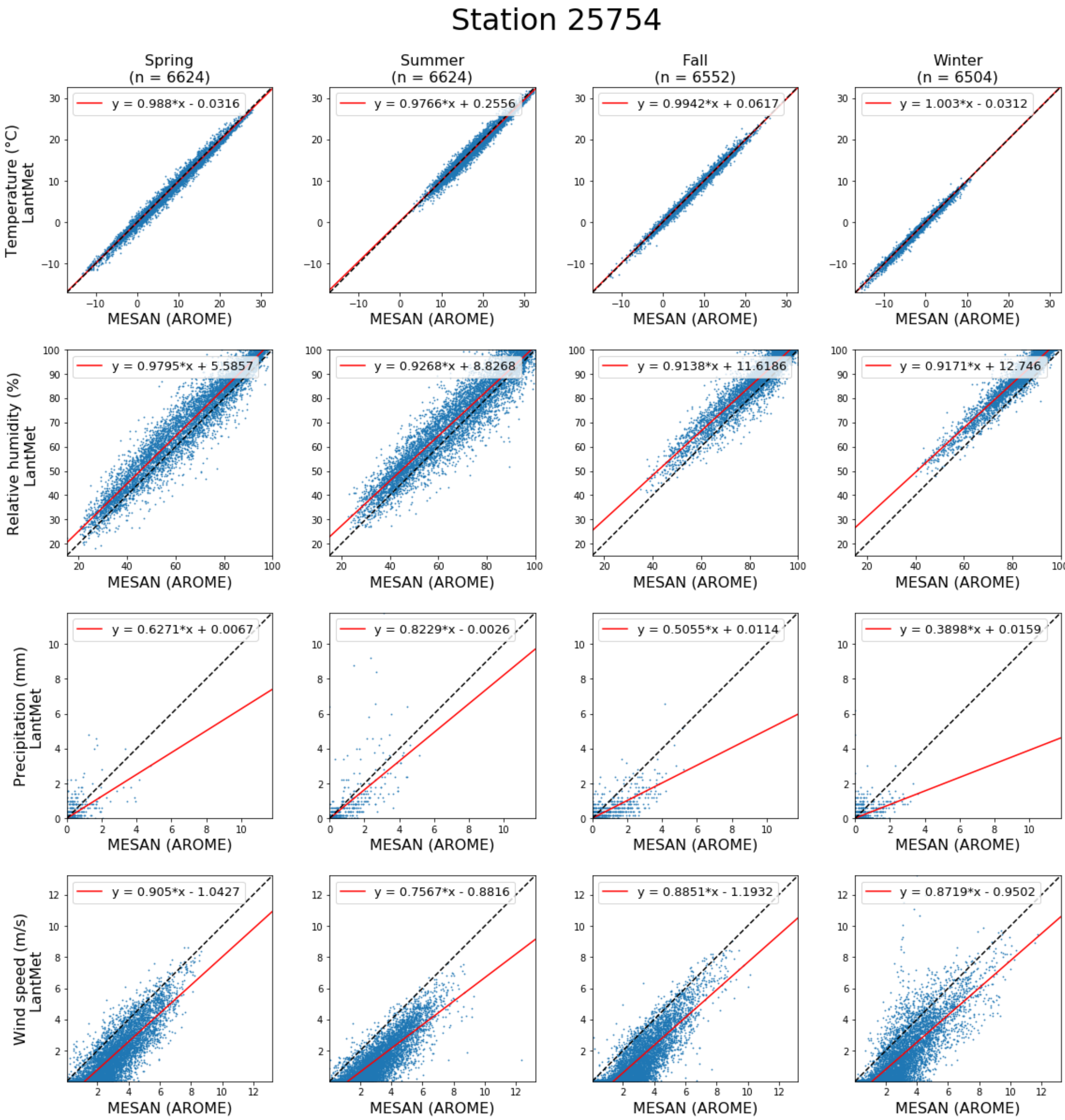

Figure 13: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

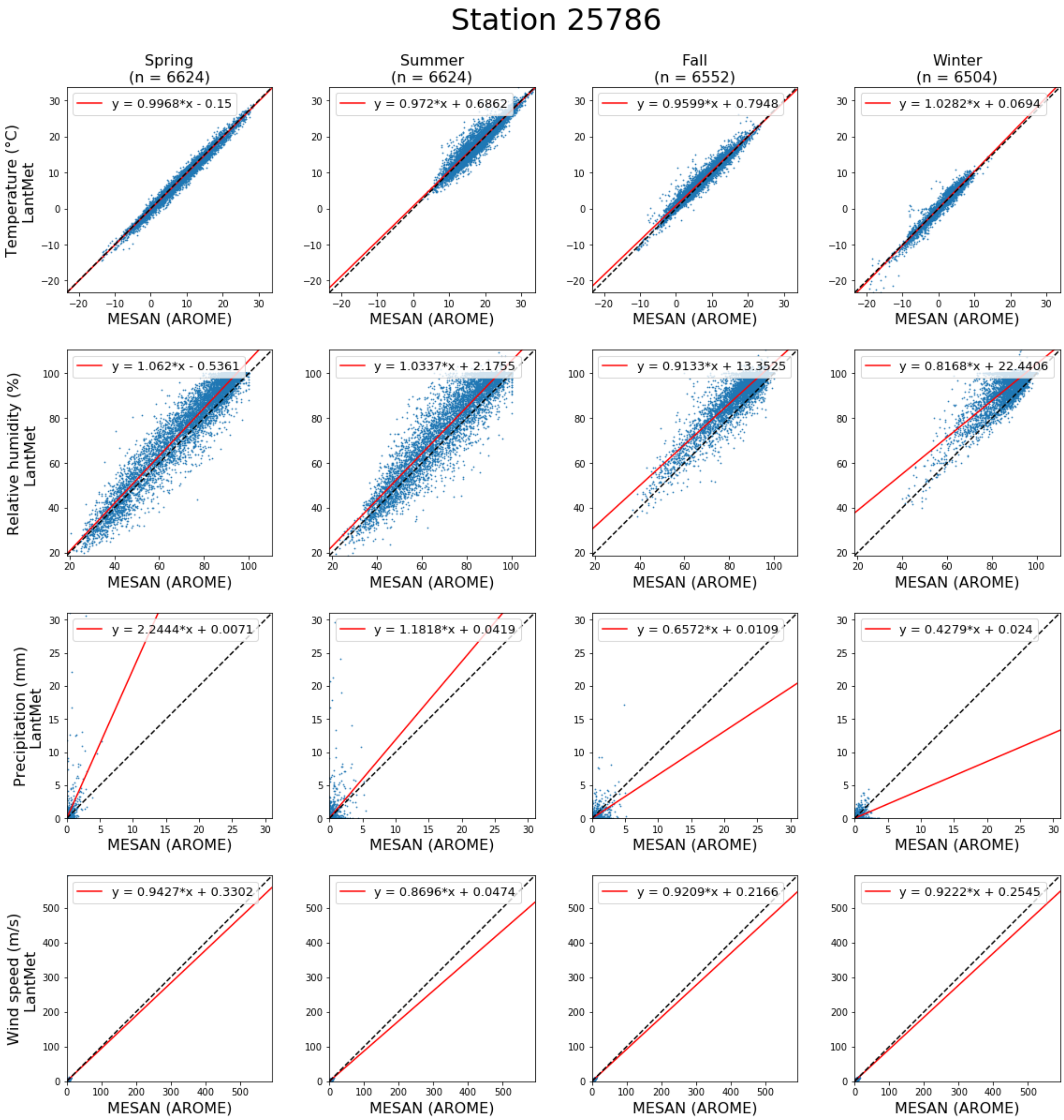

Figure 14: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

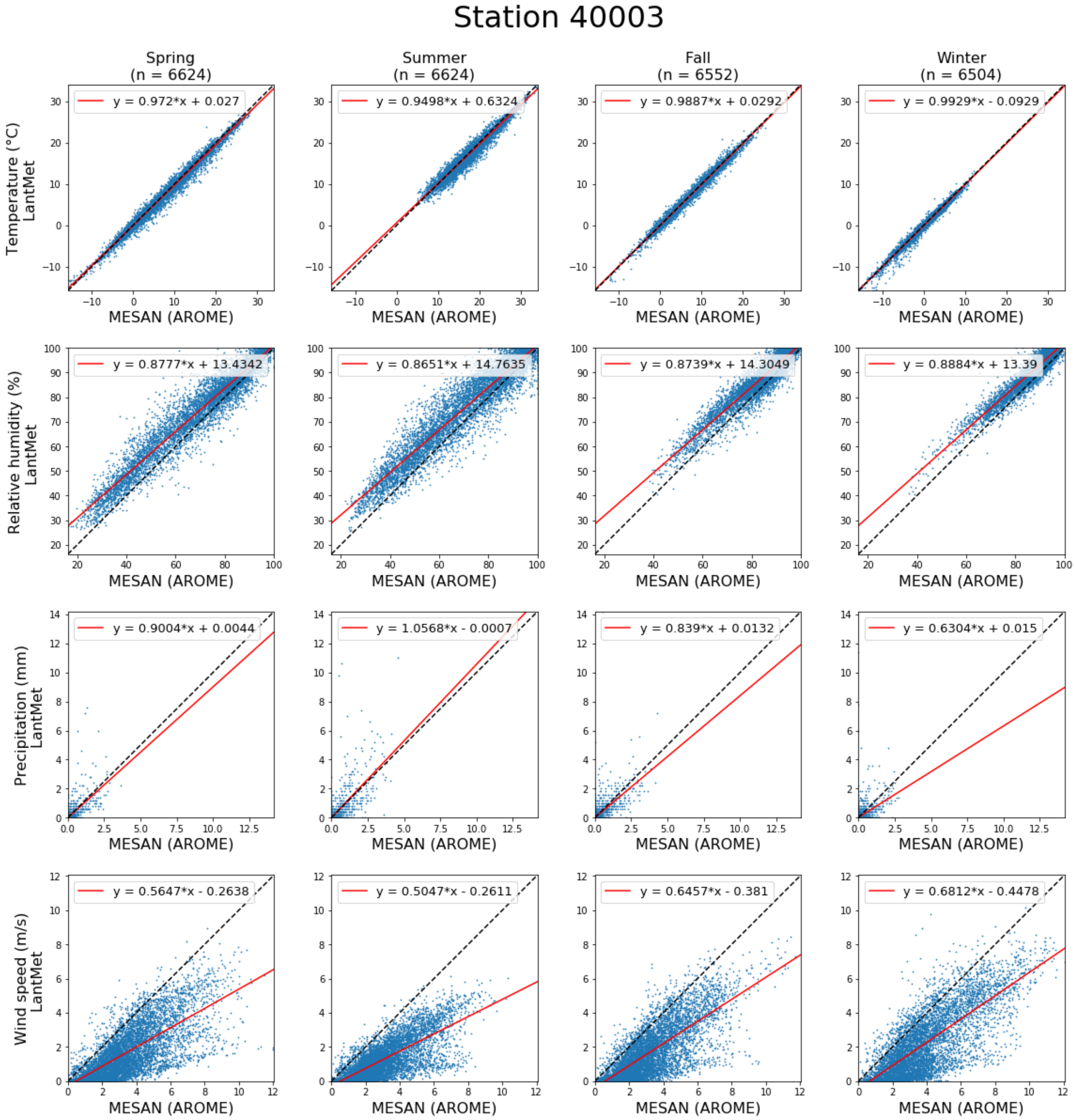

Figure 15: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

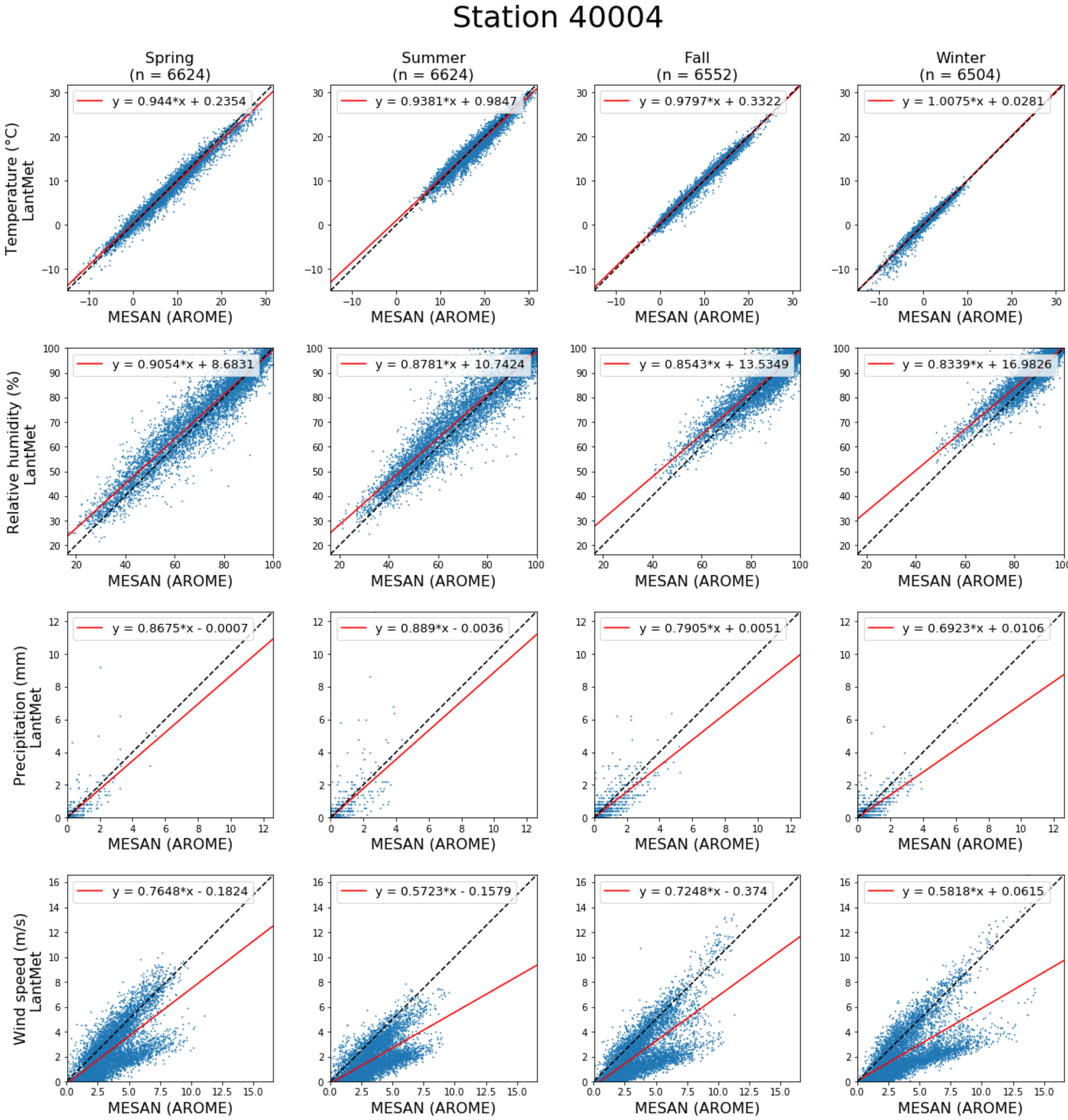

Figure 16: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

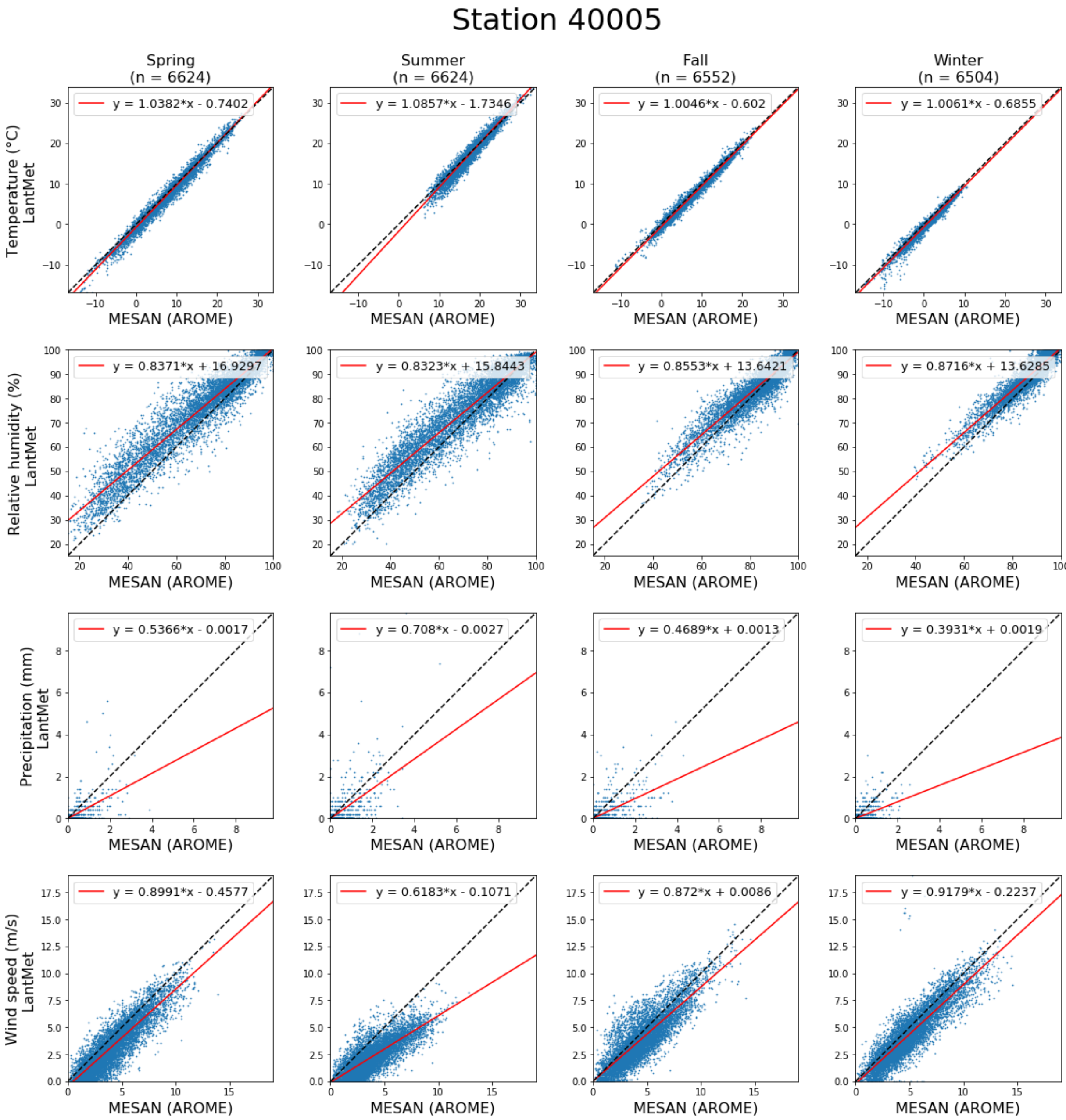

Figure 17: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

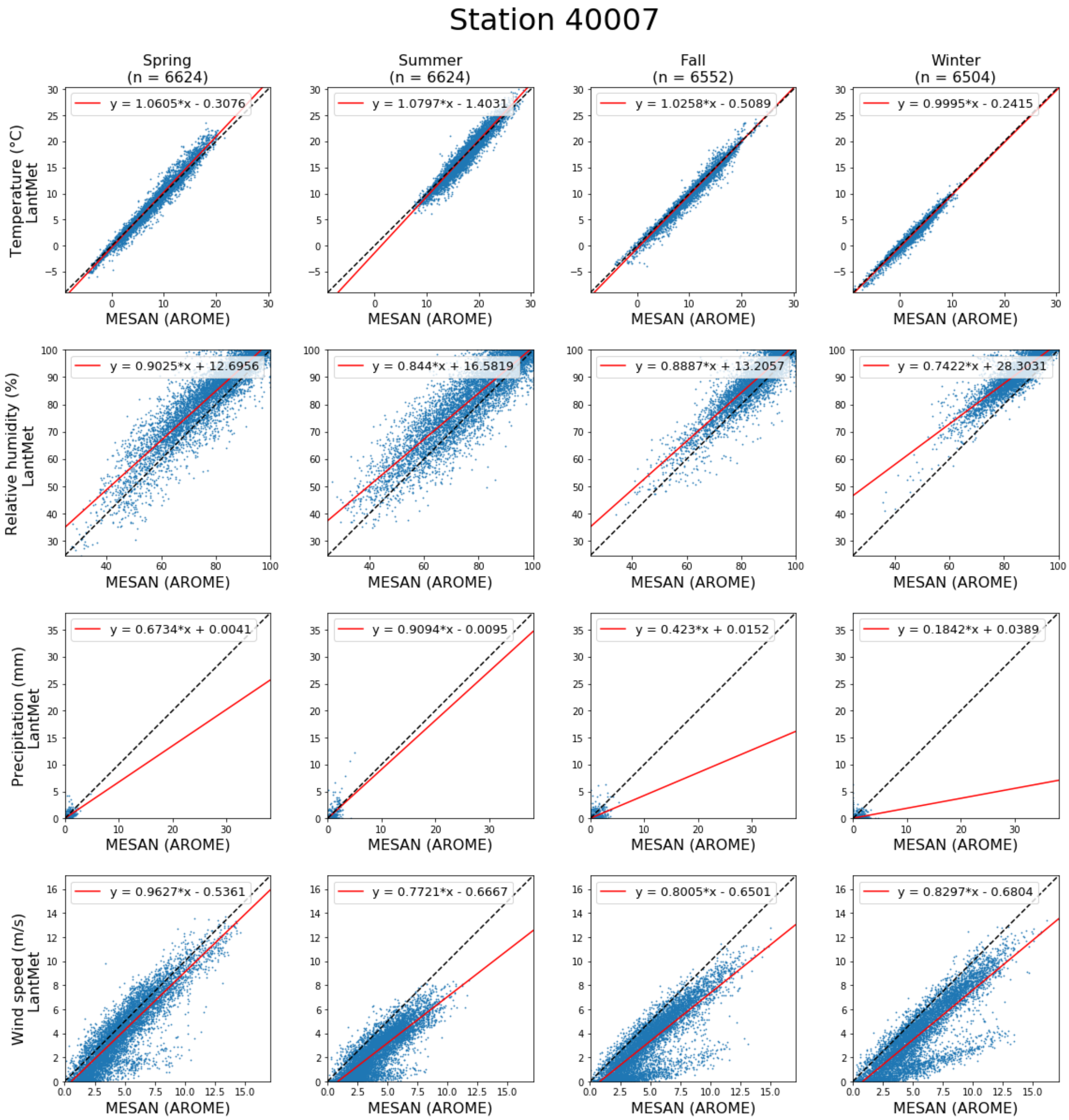

Figure 18: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.

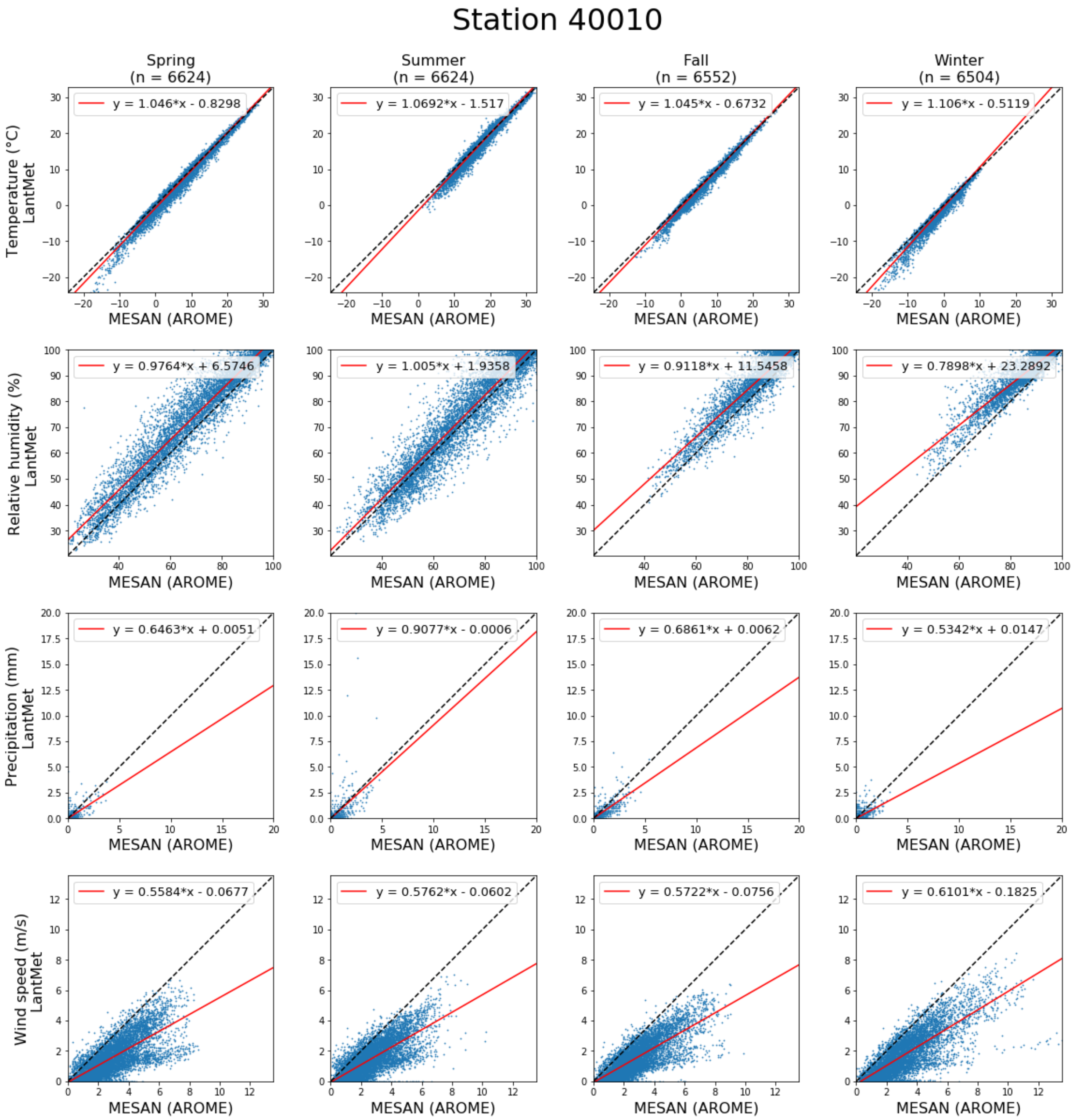

Figure 19: Scatter plots of interpolated data from the MESAN(AROME) dataset and observed data from the LantMet dataset for every hour between the first of March 2017 and the last of February 2020. The red lines visualizes a linear regression model fitted to the data while the dashed black line is a visualization of the ideal case where the interpolated and observed data match completely.