Joint modeling of low temperature and low wind speed events over Europe conditioned on winter weather regimes

Paulina Tedesco

Supervisors: Jana Sillmann, Alex Lenkoski, Frode Stordal

Thesis submitted for the degree of

Master of Science in Meteorology and Oceanography

60 Credits

Department of Geosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

September 15, 2020

Supervisors: Jana Sillmann, Alex Lenkoski, Frode Stordal

Joint modeling of low temperature and low wind speed events over Europe conditioned on winter weather regimes

This work is published digitally through DUO – Digitale Utgivelser ved UiO

`http://www.duo.uio.no/`Printed: Reprosentralen, University of Oslo

Abstract

A transition to renewable energy is needed to mitigate climate change. This transition has been led by wind energy, and it is expected to continue to be the largest source of renewable energy through to 2030 (Sawyer et al., 2017). Both energy demand and production are sensitive to meteorological conditions and atmospheric variability at multiple time scales. To accomplish the required balance between these two variables, critical conditions of high demand and low wind energy supply must be considered in the design of energy systems. The aim of this thesis is twofold. Firstly, investigate the impacts of large-scale weather regimes on cold and weak wind events during the extended boreal winter season (NDJFM). Secondly, to establish a methodology for modeling the joint distributions without making any assumptions about the marginal distributions.

The analysis of 38 years of hourly high-resolution ERA5 reanalysis data proves that the weather regimes are important predictors for both low temperature and low wind speed events over Europe. Blocking conditions, such as those observed during the Negative Phase of the North Atlantic Oscillation and Scandinavian Blocking, are associated with cold and weak wind events. Compound events are observed more than 10% of the days over large geographical areas during blocking conditions. Nevertheless, high probabilities are also observed during the Atlantic Ridge, and to some extent, during the Positive Phase of the North Atlantic Oscillation. Dependency between cold events and weak wind events is proved to be statistically significant. The correlations between the events are higher when computed for each month separately compared to the entire winter season, revealing a strong seasonality. The highest correlations values are associated with the Negative Phase of the North Atlantic Oscillation, $\rho = 0.84$, but values as high as 0.7 are registered for all the regimes. A methodology for modeling the bivariate joint distributions of low temperature and low wind speed events is described. In this context, the concept of Gaussian copulas is used to mathematically model the correlated nature among them. The marginal distributions are modeled with logistic regressions defining two sets of binary variables as predictors, the weather regimes and the months of the extended winter season.

Acknowledgements

I would like to express my sincere gratitude to my supervisors: Jana Sillmann (CICERO), Alex Lenkoski (Norwegian Computing Center), and Frode Stordal (University of Oslo). Their guidance and feedback helped me in all the time of research and writing of this thesis.

I would also like to thank Hannah C. Bloomfield (Department of Meteorology, University of Reading, Reading, UK) for assistance with the computation of the daily weather regimes.

Finally, a special thanks to my family: Agustín, Valentin, and Ainhoa, for their patience and encouragement.

List of Figures

2.1 NAO+ loading patterns for January, April, July, and October.The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 7

2.2 Maps showing correlation during 1950-2000 between the NAO+ index and monthly surface temperature departures for the three months centered on the month of interest. For example, the January pattern shows the correlation between the January values of the teleconnection index and the monthly temperature departures during December, January, and February. CPC, 2012. 8

2.3 As Figure 2.1 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 9

2.4 As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 10

2.5 As figure 2.1 but for SCAND positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 10

2.6 As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 11

2.7 As Figure 2.1 but for EA/WR. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. . . 11

2.8 As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012. 12

2.9 Density and scatterplot of a Bivariate Gaussian Distribution. The density of the joint distribution is obtained by joining a Gaussian Copula ($\rho = 0.5$) with two identical standard Gaussian univariate distributions. Sample size: $n = 275$. Gräler, n.d. 18

2.10 Density distribution and scatterplot of Gaussian copula ($\rho = 0.5$), with uniform marginal distributions. Sample size: $n = 275$. Gräler, n.d. 19

3.1 ERA5 minimum daily temperature climatology at 2 meters height. Figure based on ERA5data (NDJFM, 1979–2017). 26

3.2 ERA5 maximum daily wind climatology at 10 meters height. Colours show maximum daily wind speeds. Figure based on ERA5data (NDJFM, 1979–2017). 26

3.3 Four regimes of atmospheric circulation in the North Atlantic-European domain,AR, SCAND, NAO-, NAO+. Colours show the Z500 anomaly (m). Area of study: 27N-81N, 95.5W-45E. Figure based on ERA5 data (NDJFM, 1979–2017). 28

4.1 Mean meteorological surface impacts of the four WRs. Colours show maximum daily temperature anomalies ($^{\circ}C$). Area of study: 20N-80N, 90W-60E. Figure based on ERA5 data (NDJFM, 1979–2017). 31

4.2 Mean meteorological surface impacts of the four WRs. Colours show minimum daily wind speed anomalies (m/s). Area of study: 20N-80N, 90W-60E. Figure based on ERA5 data (NDJFM, 1979–2017). 32

4.3 Total number of temperature observations by regime. Blue columns represent observations above the 10th percentile computed for the winter season (NDJFM) on a grid level, and orange columns represent observations below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 33

4.4 As Figure 4.1 but for the wind speed variable. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 35

4.5 Total number of joint temperature and wind speed events by regime over the European domain. The threshold is computed on a grid level for each variable separately as the 10^{th} percentile considering all the days in the winter season (NDJFM). Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 36

4.6 Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 37

4.7 Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 38

4.8 Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile for each WR. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 39

4.9 Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily maximum wind speed below the 10th percentile for each WR. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 40

4.10 Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). . . . 41

4.11 Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily maximum wind speed the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 43

4.12 Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). . . . 45

4.13 Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 46

4.14 Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 48

4.15 Seasonality of difference between the joint probabilities and the product of the marginal probabilities. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).). 50

4.16 Difference between joint probabilities and product of marginals for each WR. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 51

4.17 Seasonal difference between joint probabilities and product of marginals for each WR. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 52

4.18 Correlations (ρ) between low temperature events and low wind speed events for the four WRs. The threshold is defined as the below the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 53

4.19 Seasonality of correlations (ρ) between low temperature events and low wind speed events for the four WRs. The threshold is defined as the below the 10^{th} percentile. Locations with no low temperature event during the selected period have no colour. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 54

4.20 Significant BSS for minimum daily surface temperature for each WRs. A monthly seasonal model was used as a reference. Colours show significant BSS values at the 10% level that are not zero, and the color bar is truncated at $BSS = 0$ and $BSS = 0.7$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 56

4.21 As figure 4.20 but for maximum daily 10m wind speeds. Colours show significant BSS values at the 10% level, and the color bar is truncated at $BSS = 0$ and $BSS = 0.3$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 57

4.22 Seasonality of ignificant BSS for minimum daily surface temperature for each WRs. A monthly seasonal model was used as a reference. Colours show significant BSS values at the 10% level that are not zero, and the color bar is truncated at $BSS = 0$ and $BSS = 0.3$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 58

4.23 As figure 4.22 but for maximum daily 10m wind speeds. Colours show significant BSS values at the 10% level, and the color bar is truncated at $BSS = 0$ and $BSS = 0.3$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 59

4.24 Seasonality of joint probabilities from the copula model for each WR. Colours show the probabilities of occurrence of low temperature and low wind speed events computed with the copula model for minimum temperatures and daily maximum wind speeds below the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 61

4.25 Significant BSS at the 10% level for joint probabilities modeled with Gaussian copulas and logistic regressions, aggregated by WRs. An independent model, with no correlation between the temperature and the wind speed events, was employed as the reference model. Colours show significant BSS values that are not zero, and the color bar is truncated at $BSS = 0.03$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 62

4.26 Significant BSS for joint probabilities modeled with Gaussian copulas and logistic regressions, aggregated by WRs (AR, SCAND, NAO-)and month (DJF). An independent model, with no correlation between the temperature and the wind speed events, was employed as the reference model. Colours show significant BSS values that are not zero, and the color bar is truncated at $BSS = 0.03$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 63

C.1 Seasonality of coefficients of modeled minimum daily surface temperature below the 10^{th} percentile. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = 0$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 85

C.2 As C.1 but for maximum daily wind speed at 10 meters. Colours show coefficients, β, and the colorbar is truncated at $\beta = -0.1$ and $\beta = -0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 85

C.3 Coefficients of modeled minimum daily surface temperature below the 10th percentile for each WR. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = 0$ and $\beta = 0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 86

C.4 As C.3 but for maximum daily wind speed at 10 meters. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = -0.1$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 87

C.5 Coefficients of modeled minimum daily surface temperature below the 10^{th} percentile for each month and WRs. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = 0$ and $\beta = 0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 88

C.6 As C.5 but for maximum daily wind speed at 10 meters. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = -0.1$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 89

C.7 Seasonality of modeled probabilities of occurrence of minimum daily surface temperature below the 10^{th} percentile for each month. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 89

C.8 As C.7 but for maximum daily wind speed at 10 meters. Colours show probabilities, p, and the colorbar is truncated at $p = 0.35$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 90

C.9 Modeled probabilities of occurrence of minimum daily surface temperature below the 10th percentile for each WR. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 90

C.10 As C.9 but for maximum daily wind speed at 10 meters. Colours show probabilities, p, and the colorbar is truncated at $p = 0.35$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 91

C.11 Modeled probabilities of occurrence of minimum daily surface temperature below the 10^{th} percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 92

C.12 As C.11 but for maximum daily 10m wind speeds. Colours show probabilities, p, and the color bar is truncated at $p = 0.30$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 93

C.13 Modeled probabilities of occurrence of minimum daily surface temperature below the 15th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 95

C.14 Modeled probabilities of occurrence of minimum daily surface temperature below the 5th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 96

C.15 Modeled probabilities of occurrence of maximum daily 10m wind speeds below the 15th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 97

C.16 Modeled probabilities of occurrence of maximum daily 10m wind speeds below the 5th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 98

C.17 99

C.18 Independence between low temperature an low wind speed events by regime. Colours show p-values obtained with the chi-square test. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 99

C.19 100

C.20 Independence between low temperature an low wind speed events by regime and month. Colours show p-values obtained with the chi-square test. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 100

C.21 Joint probabilities from copula model for each . Colours show the probabilities of occurrence of low temperature and low wind speed events computed with the copula model for minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017). 101

List of Tables

4.1 Proportions of events in each WR computed over the entire European domain. The first column, $\%t > q$, represents the percentage of the total number of days with temperatures above the threshold over the entire grid, computed with respect to the total number of days in each cluster. Analogously, the second column, $\%t < q$, represents the percentage of the total number of events in Europe. The ratio between these two columns is written in the third column. The last column is the sum of the values in the first two columns and represents the percentage of the number of days classified in each WR. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 34

4.2 As Tanble 4.1 but for wind speed events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 35

4.3 As Table 4.1 but for joint temperature and wind speed events. The threshold for defining cold events and weak wind events is the 10^{th} percentile of each variable computed for all the days in the winter season (NDJFM). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 36

4.4 Maximum empirical marginal probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 42

4.5 Mean empirical marginal probabilities of low temperature events p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). . 42

4.6 Maximum empirical marginal probabilities of low wind speed events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 44

4.7 Mean empirical marginal probabilities of low wind speed events (p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). . 44

4.8 Maximum empirical joint probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). . 47

4.9 Mean empirical joint probabilities of low temperature events (p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). . 49

4.10 Maximum (ρ_X), minimum (ρ_N), and average (ρ_μ) correlations for each regime. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 52

4.11 Maximum significant BSS for low temperature events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 58

4.12 Maximum significant BSS for low wind speed events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 60

B.1 Python dependencies. 84

C.1 Maximum probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 91

C.2 Maximum probabilities of low wind speed events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017). 94

Acronyms

AR Atlantic Ridge. i, v, ix, 2, 5, 6, 13, 14, 27–29, 32, 34, 35, 38–40, 42, 43, 45, 47, 48, 52, 53, 55–57, 60, 63, 65, 66, 68, 72, 92–94, 97

BS Brier Score. 21, 22, 30, 68

BSS Brier Skill Score. viii, ix, xiv, 22, 29, 30, 55–60, 62, 63, 68, 72

CDF cumulative density function. 16–19

CPC Climate Prediction Center. 5, 6

EA East Atlantic Pattern. iv, v, 2, 5–7, 9–12

EA/WR East-Atlantic West Russia Pattern. iv, 6, 8, 11

EATs Euro-Atlantic teleconnections. 6

ECMWF European Centre for Medium-Range Weather Forecasts. 25

EOFs Empirical Orthogonal Functions. 5, 27

IFS Integrated Forecasting System. 25

MCMC Markov chain Monte Carlo. 30, 72

MSLP Mean Sea Level Pressure. 5

NAO North Atlantic Oscillation. 2, 5–7, 13, 31, 32, 34, 65

NAO+ Positive Phase of the North Atlantic Oscillation. i, iv, v, 5–8, 13, 27–29, 31–36, 39–43, 45, 47, 49, 53, 55–57, 60, 65, 66, 71, 72, 93, 94, 96

NAO- Negative Phase of the North Atlantic Oscillation. i, v, ix, 5, 13, 14, 27–29, 31, 32, 34, 35, 38–40, 42–45, 47, 48, 50, 53, 55–57, 60, 63, 65–68, 71, 72, 88, 92–94, 96

NCEP National Senters for Environmental Prediction. 5

NetCDF Network Common Data Form. 25, 83

PCs Principle Component. 27

REOFs Rotated Empirical Orthogonal Functions. 5

SCAND Scandinavian Blocking. i, iv, v, ix, 2, 6, 8, 10, 13, 14, 27–29, 31, 32, 34, 35, 39, 40, 42–44, 47, 48, 53, 55–57, 60, 63, 65–68, 71, 88, 92–94, 96

SDGs Sustainable Development Goals. 1

TLp Teleconnection Patterns. iv, v, 5–7, 9–13

WRs Weather Regimes. v, vii–xi, 2, 3, 5, 9, 12, 13, 18, 26, 28, 29, 31, 32, 34, 38, 40, 41, 45, 46, 49–51, 53–59, 62, 63, 65–69, 71, 72, 82, 88, 92, 95–98

Z500 geopotential height at 500 hPa. v, 12, 26–28, 65

1 Introduction

1.1 Motivation and Background

Affordable and clean energy is one of the UN Sustainable Development Goals (SDGs). Energy is also crucial for achieving almost all of the other SDGs. The energy sector currently accounts for more than two-thirds of the global greenhouse gas emissions. Consequently, a change of the entire energy sector is required to meet the SDGs and the Paris agreement targets (Sawyer et al., 2017). In particular, a rapid decarbonization of the global energy system is needed to limit global warming to a maximum of 1.5 $^{\circ}C$ (Rogelj et al., 2015), by increasing the share of wind and solar power generation of the total power generation. Europe is leading this transition, although it is still one of the world's biggest energy consumers and greenhouse gas emitters (Liobikienė and Butkus, 2017).

The production of clean energy is highly weather-dependent; for instance, wind power prodction depends on the wind speed. Meanwhile, the demand tends to increase when the weather is cold, and electricity is needed for heating, or when it gets too hot. Therefore, the expansion of renewable energy over Europe increases the sensitivity of power systems to meteorological conditions and variability (H C Bloomfield et al., 2016; H C Bloomfield et al., 2018; Wohland et al., 2017; Collins et al., 2018; Zeyringer et al., 2018; Wiel, Hannah C Bloomfield, et al., 2019 **"a"**), complicating the energy transition. Given that electricity production and demand depend on the weather, they vary on multiple timescales (e.g. Sinden, 2007; Bessec and Fouquau, 2008; H C Bloomfield et al., 2016). Hence, it is important to understand how large-scale circulation systems influence peaks of demand and energy production to identify periods of over- and under-supply. It has been shown that large scale anticyclonic pressure systems over Europe can lead to low wind and solar production and high demand, resulting in energy shortfall (H C Bloomfield et al., 2018; Wiel, Hannah C Bloomfield, et al., 2019). Further studies are needed to understand the impact of atmospheric variability on surface variables that simultaneously affect energy production and demand, e.g., cold spells and low wind conditions.

Extreme weather conditions can lead to high electricity demand and low renewable power production. How to deal with periods of low production is a big challenge in the design of renewable energy systems. (Huber, Dimkova, and Hamacher, 2014. Subseasonal and seasonal

forecasts help improve decision-making and planning. They are of value for power producers, to better prepare for extreme meteorological events, and could be an important planning tool for traders, plant operators, and investors for managing climate variability related risk (Cortesi et al., 2019), as they provide relevant information for price forecasting (Pinson, 2013) **check**). But high energy shortfall events are a combination of multiple drivers that contribute to societal and environmental risk, and studies in this field have traditionally focused on single drivers without considering interactions between them. Low energy production combined with high demand due to electric heating requirements leads to high energy shortfall from November to March. Meteorological conditions preceding high energy shortfall events - posing a risk for European energy security - are described as anomalous high pressure systems combined with below normal temperatures (Wiel, Stoop, et al., 2019). Nevertheless, as far as we know, no previous research has proposed a meteorological based methodology to model the multivariate probabilities of these events, allowing for better preparation.

A variety of indices have been constructed to describe European climate variability through daily synoptic-scale weather patterns. A method that has been proven to be useful in weather forecasting and climate change applications is the computation of Weather Regimes with the k-means algorithm (e.g. Neal et al., 2016, Ferranti, Corti, and Janousek, 2015; Neal et al., 2016; Matsueda and Palmer, 2018). In this thesis, four weather regime patterns in the Euro-Atlantic region where derived from geopotential height in the mid-troposphere during the boreal winter using the k-means algorithm (Michelangeli, Robert Vautard, and Legras, 1995; Cassou, 2008). The impact of the (WRs) at the surface is also relevant for energy applications (Wiel, Hannah C Bloomfield, et al., 2019. The four weather regime patterns derived with this methodology can be related to the negative phase of the East Atlantic Pattern (AR), the positive phase of the Scandinavian Blocking (SCAND) pattern, and both phases of the North Atlantic Oscillation (NAO). The response of power systems to these patterns across the extended European region has recently been studied, with a focus on the NAO (e.g., (Ely et al., 2013; Thornton et al., 2017; H C Bloomfield et al., 2018, David James Brayshaw et al., 2011; Zubiate et al., 2017; H C Bloomfield et al., 2018). Studies show that WRs are useful for subseasonal to seasonal energy applications (Cassou, 2010; Wiel, Hannah C Bloomfield, et al., 2019).

Modeling the dependence between demand and energy production is of key importance to understand the occurrence of energy shortfall and prevent it by redesigning the energy systems. However, modeling multivariate distributions can be a challenge. Copulas offer a powerful and

flexible tool to model dependency between different variables. They return the joint probability of events as a function of the marginal probabilities of each event. This makes copulas attractive, as the univariate marginal behavior of random variables can be modeled separately from their dependence. The semi-parametric Bayesian Gaussian copula methods used here estimate multivariate relationships between variables with univariate marginal distributions that cannot be well approximated with a simple parametric model (D. Hoff, 2007), as is the case of our event variables, making it a promising candidate for modeling the joint probabilities of low temperature and low wind speed events.

This thesis aims to propose a new methodology for modeling meteorological compound events associated with high energy shortfall. We show that the WRs provide useful information for predicting marginal probabilities of low temperatures and low wind speeds and that these events are correlated. Gaussian copulas allow us to model the dependency between low temperature and low wind-speed events as a function of their marginals, separately from their dependence. The high spatial resolution of hourly ERA5 reanalysis is exploited to estimate joint probabilities across Europe on a $0.25^\circ \times 0.25^\circ$ grid.

The remainder of this thesis is organized as follows. An overview of the theories and concepts relevant to the topic of this thesis is briefly introduced in section 2. The data and the methodology developed to compute the probabilities of low wind speed and temperature events are described in section 3. Results are presented in section 4 and discussed in section 5. Finally, 6, provides the conclusions. Additional figures and supporting material are available in Appendix 8.

2 Theory

2.1 Euro-Atlantic teleconnections and Weather Regimes

The atmospheric circulation is well-known for its variability in multiple time scales being reflected in weather patterns and circulation systems. Thus, the knowledge of these variability patterns gives a certain amount of predictability in remote locations (Quadrelli and J. Wallace, 2004). There is a vast type of variability patterns described in the literature; some labeled as Teleconnection Patterns (TLp), oscillations, clusters, seesaws, or modes, calculated using different analysis techniques.

Atmospheric teleconnection indices are typically derived from the first few principal components of seasonal, or monthly, mean of upper-atmosphere geopotential height or MSLP, or by using techniques such as Empirical Orthogonal Functions (EOFs) or Rotated Empirical Orthogonal Functions (REOFs). These are usually employed to assess European climate variability (Barnston and Livezey, 1987). The NCEP's Climate Prediction Center (CPC) publishes regularly the indices computed with the REOFs method, which has gained popularity among climate scientists. However, the associated TLp are only made available as images, which prevents computing the indices from seasonal predictions or other reanalysis datasets. Another limitation of this methodology to assess climate variability is that it imposes symmetric variability patterns. Instead, classifying WRs on timescales of days by using the k-means algorithm provides an alternative description of variability that does not enforce identical oscillation phases. (Cassou, 2008).

The clustered WRs and the EOFs refer, both, to the recurring and persistent large-scale circulation patterns (J. M. Wallace and Gutzler, 1981). A key difference is that the patterns computed from the EOFs typically last for several weeks or months and span vaster geographical area than the WRs (Thompson and J. M. Wallace, 1998), whereas the WRs vary on time-scales of days (Cassou, 2008). Meanwhile, the four clusters obtained resemble some of the phases of the main TLp computed by the CPC, and are therefore discussed in this section. Two of them are consistent with the spatial patterns of the two opposite phases of the North Atlantic Oscillation (NAO), and are therefore called NAO+ and NAO- regimes. The third regime is named Atlantic Ridge (AR) and resembles the negative phase of the East Atlantic Pattern (EA),

which is represented by a positive anomaly over the Atlantic Ocean and a negative anomaly over Scandinavia. The fourth is called Scandinavian Blocking (SCAND), and is characterized by a strong anomalous height anomaly over Northern Europe and a weaker negative anomaly over the Atlantic Ocean, similar to the atmospheric flow during blocking events in Europe.

2.1.1 Euro-Atlantic teleconnections (EATs)

TLp are often defined as recurrent and persistent large-scale structures with centers of opposite sign, that vary in intensity and position at seasonal, inter-annual, and decadal time periods (J. M. Wallace and Gutzler, 1981; Trenberth et al., 1998; Quadrelli and J. Wallace, 2004). They are characterized by correlations between geopotential heights on a given pressure level at widely separated locations, occurring simultaneously, and affecting the atmospheric circulation variability. These preferred modes of low-frequency typically last for several weeks to months, but can sometimes be prominent for several consecutive years.

TLp are associated with large-scale changes in the atmospheric wave and jet stream patterns, and influence temperature, rainfall, storm tracks, and jet stream location/intensity over vast areas. The four most prominent patterns over Europe exist over all months and are denominated (a) North Atlantic Oscillation (NAO), (b) East Atlantic Pattern (EA), (c) Scandinavian Blocking (SCAND), and (d) Atlantic Ridge (AR), also called East-Atlantic West Russia Pattern (EA/WR). These patterns imprint different temperature, wind speed, and precipitation conditions that can be associated with impacts on the European energy system (David James Brayshaw et al., 2011; Cradden et al., 2017; Zubiate et al., 2017). The following discussion relies on information published by the CPC.

(a) NAO

One of the TLp that explains most of the atmospheric variability in all seasons is the NAO (Barnston and Livezey, 1987), consisting of a north-south dipole of anomalies. The NAO index is based on the surface sea-level pressure difference between the Subtropical (Azores) High and the Subpolar Low.

The positive phase of the NAO corresponds to below-normal heights and pressure in high latitudes in the North Atlantic and above-normal heights and pressure over the Central North Atlantic, the Eastern United States and Western Europe (see Figure 2.1), whereas the negative phase corresponds to the opposite patterns in these regions.

Strong NAO+ conditions are associated with above-normal temperatures in Northern Europe and below-normal temperatures in Southern Europe (see Figure 2.2). They are also associated

with above-normal precipitation over Northern Europe and Scandinavia and below-normal precipitation over Southern and Central Europe. Opposite patterns of temperature and precipitation anomalies are typically observed during strong negative phases of the NAO.

This TLp presents notably intraseasonal and interannual variability, and it is common to have prolonged periods (several months) of both positive and negative phases of the pattern.

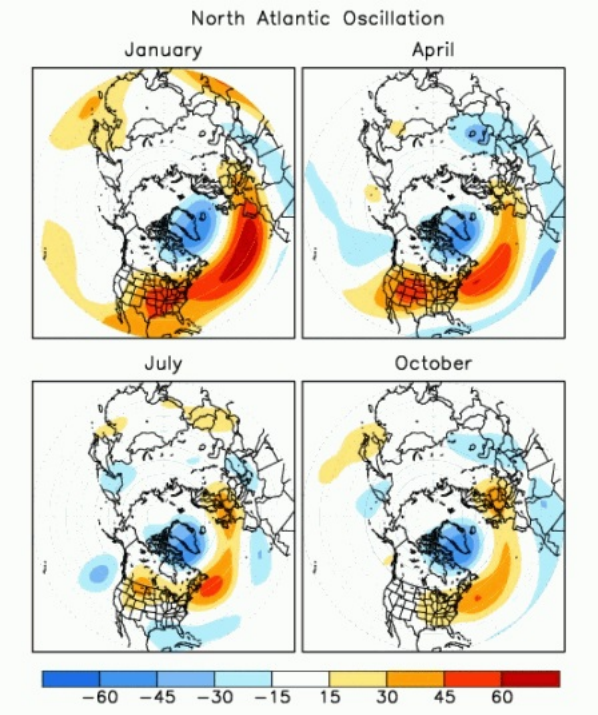

Figure 2.1: NAO+ loading patterns for January, April, July, and October.The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

(b) EA

The EA pattern consists of a north-south dipole of anomaly centers spanning the North Atlantic from east to west, displaced southeastward with respect to the NAO centers (see Figure 2.3). These lower-latitude centers contain a strong subtropical link associated with modulations in the subtropical ridge intensity and location, making it different from the NAO.

During a positive phase of the EA, above-average surface temperatures in Europe in all months are expected (see Figure 2.4). This phase is also associated with above-average precipitation over Northern Europe and Scandinavia, and with below-average precipitation across Southern Europe.

The EA exhibits very strong multi-decadal variability. The index shows a negative phase prevailing during 1950-1976, and a positive phase, particularly strong and persistent, prevailing

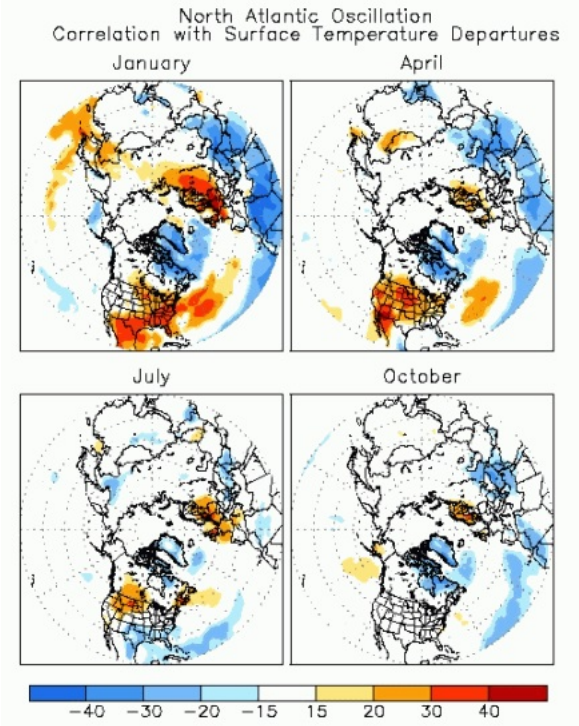

Figure 2.2: Maps showing correlation during 1950-2000 between the NAO+ index and monthly surface temperature departures for the three months centered on the month of interest. For example, the January pattern shows the correlation between the January values of the teleconnection index and the monthly temperature departures during December, January, and February. CPC, 2012.

from 1977.

(c) SCAND

The SCAND consists of a primary circulation center over Scandinavia, with weaker centers of an opposite sign over Western Europe and Eastern Russia/western Mongolia (see Figure 2.5).

Positive height anomalies are typically observed during the positive phase of this pattern, sometimes reflecting the presence of major blocking anticyclones over Scandinavia and Russia. Over Western Europe, it is associated with below-average temperatures (see Figure 2.6). It is also associated with above-average precipitation across Central and Southern Europe and below-average precipitation across Scandinavia. The negative phase is associated with negative height anomalies over Scandinavia and Russia.

(d) EA/WR

The EA/WR affects Eurasia throughout the year and consists of four main anomaly centers. The positive phase is associated with positive height anomalies located over Europe and Northern China, and negative height anomalies located over the Central North Atlantic and north of the

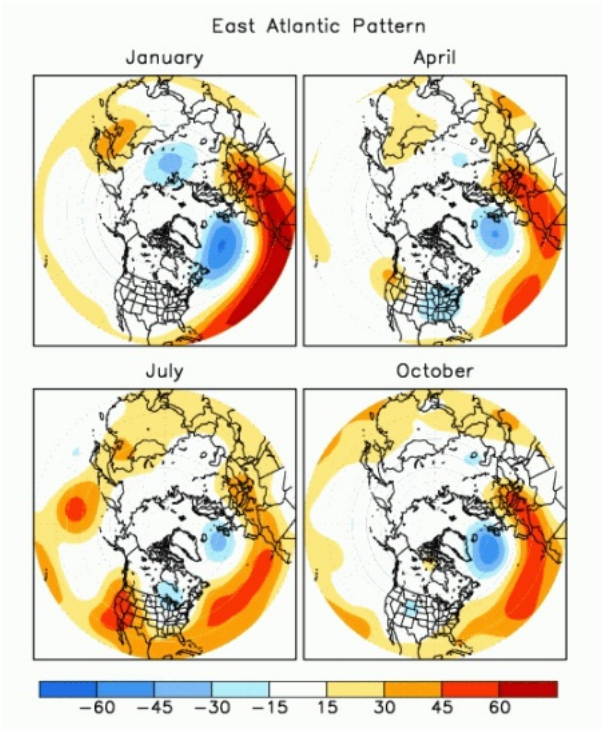

Figure 2.3: As Figure 2.1 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

Caspian Sea (see Figure 2.7).

Over Europe, it is associated with above-average surface temperature anomalies (see Figure 2.8), and below-average precipitation during the positive phase.

2.2 Weather Regimes

WRs are quasi-stationary large scale circulation patterns (Reinhold and Pierrehumbert, 1982) produced by the interaction between planetary-scale and synoptic-scale waves(Cassou, 2008), during which the character of the synoptic storms is unusually persistent (Straus, Corti, and Molteni, 2007). They typically persist for 6–10 days, are spatially well defined (typically the width of an oceanic basin), and are limited in number (Cassou, 2008). Traveling synoptic systems or storms, linked to the upper-level westerly jet stream, are embedded in these WRs. They have traditionally been used in the process of understanding the midlatitude atmospheric variability; the description of the causes of their recurrence, persistence, and transition is crucial for medium-range and seasonal-to-interannual climate prediction (Cassou, 2008; Cassou, 2010). They influence the weather at the surface hence influencing renewable power generation and electricity demand (Grams et al., 2017; Thornton et al., 2017.

2.2.

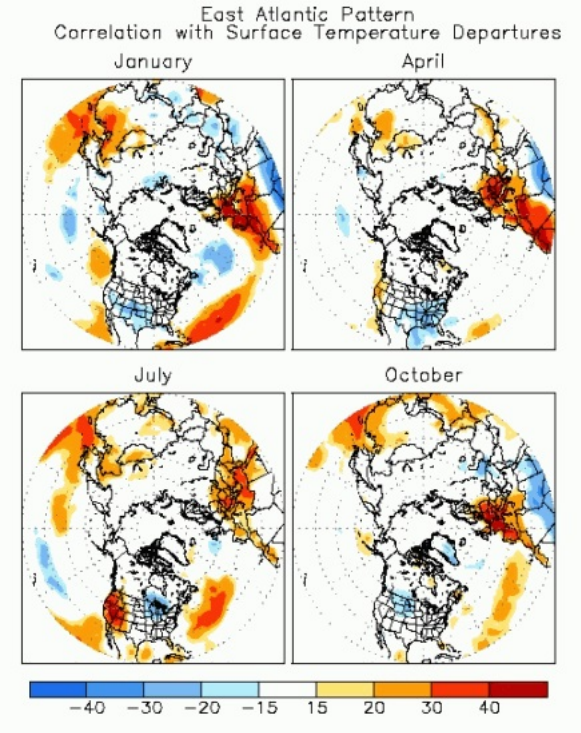

Figure 2.4: As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

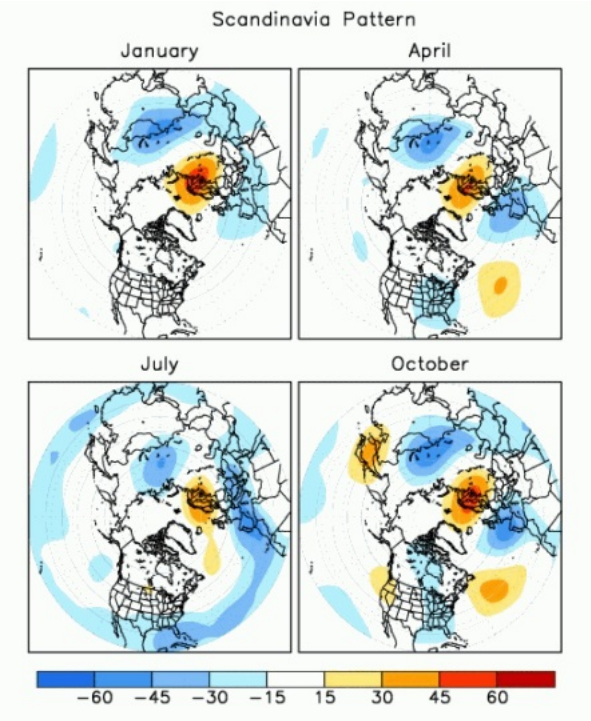

Figure 2.5: As figure 2.1 but for SCAND positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

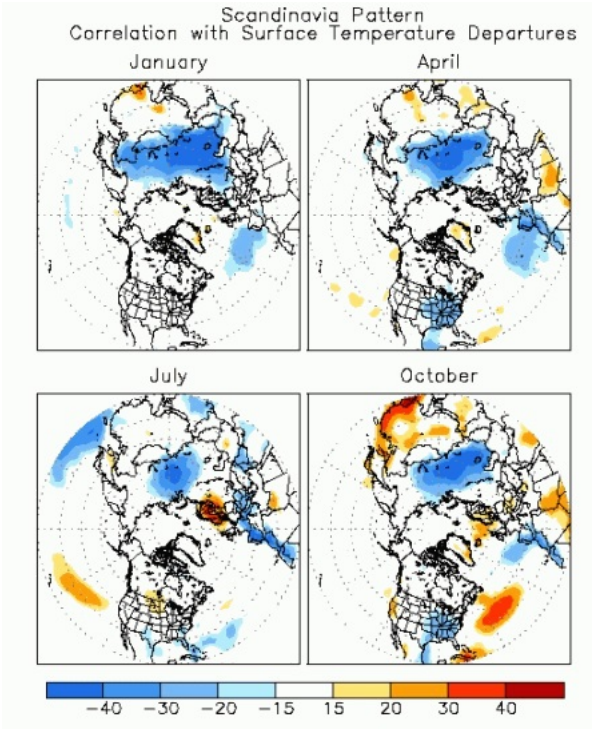

Figure 2.6: As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

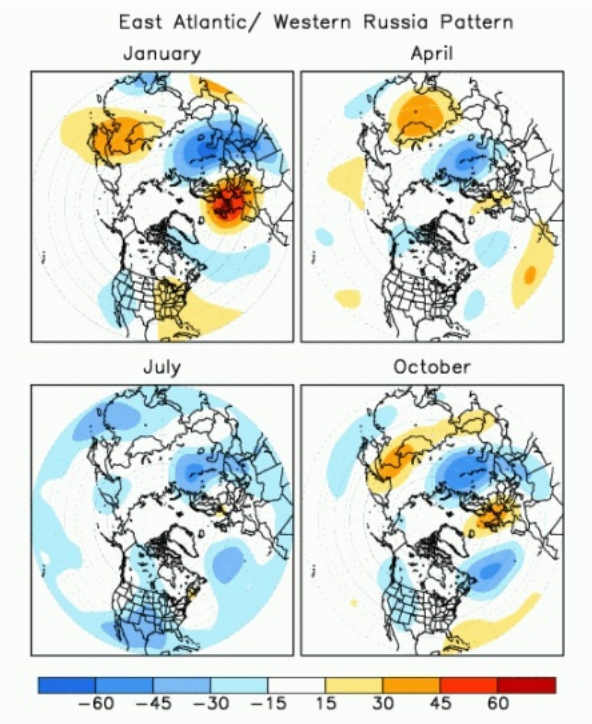

Figure 2.7: As Figure 2.1 but for EA/WR. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

2.2.

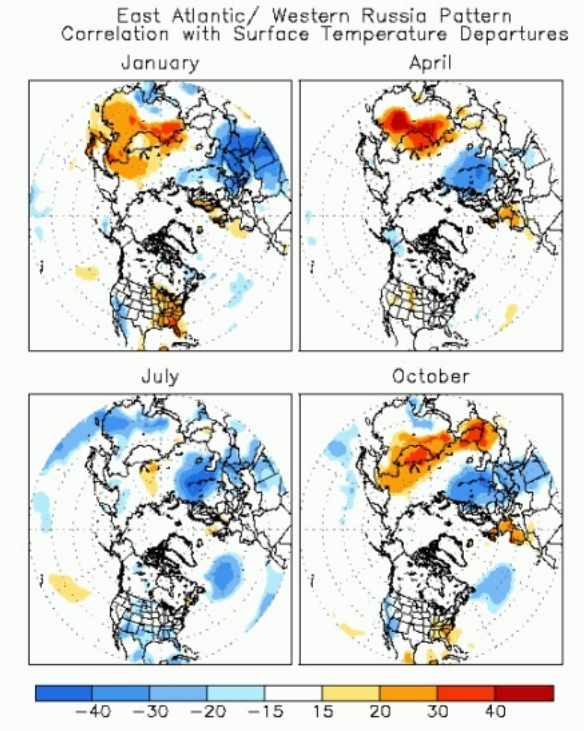

Figure 2.8: As Figure 2.2 but for EA positive. The plotted value at each grid point represents the temporal correlation between the monthly standardized height anomalies at that point and the TLp time series valid for the specified month. CPC, 2012.

WRs are traditionally obtained by using cluster analysis, more specifically, the k-means algorithm with four centroids on geopotential height at 500 hPa ((Michelangeli, Robert Vautard, and Legras, 1995; Cassou, 2008). The algorithm used to define the WRs leads to four regimes in the Euro-Atlantic area during boreal winters (November to March). It has been shown that temporal sub-sampling (Cassou, 2008) and the use of different reanalysis data (Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez, 2020; Wiel, Hannah C Bloomfield, et al., 2019) do not change the spatial structure of the regimes nor the optimal partition (k = 4). Details about the construction of the indices are provided in section 3, Methods and Data.

Considering a winter season of five months (November to March) instead of the traditional 3-month definition of the season has two advantages (Cassou, 2008). From a statistical point of view, since clustering is sensitive to sampling (Wilks, 2006), adding two months of data reinforces the significance of the WR partition. On the other hand, from a physical perspective, it fits the cycle of the Madden-Julian Oscillation consisting of the extended winter and the extended summer seasons (Wheeler and Hendon, 2004),

In agreement with R. Vautard (1990), the regimes can be interpreted as the well-established

TLp: the two first regimes correspond to the negative and positive phases of the NAO. The third regime is the Atlantic ridge, and the fourth is the Scandinavian blocking (SCAND).

Cassou (2010) shows that the probability of occurrence of extreme temperature and precipitation events can be related to the four WRs. His work indicates that NAO+ precludes and (NAO-) favours the occurrence of cold extremes over entire Europe. During the AR, the probability of cold events is significantly increased over the Iberian Peninsula, whereas during the SCAND, cold events are favoured in Central Europe extending towards France.

2.2.1 Impact of the WRs across Europe during the winter season on surface variables

The position of anomalous pressure systems and planetary waves- acting in different timescales disturbs the zonal flow at 500 hPa, which in turn influences the progression of WRs that influences surface variables of relevance for energy applications, such as temperature and wind speed. Studies confirm that, in general, blocking conditions are associated with above-average demand and below-average wind and solar generation in Central and Northern Europe (Grams et al., 2017; Wiel, Hannah C Bloomfield, et al., 2019 **a**), whilst more zonal wind conditions lead to above-average wind generation in Central–Northern Europe and below-average generation in Southern Europe (Grams et al., 2017).

(a) NAO+

It has been shown that the NAO has an influence over the energy demand (Ely et al., 2013; Thornton et al., 2017; H C Bloomfield et al., 2018) and wind power (David James Brayshaw et al., 2011; Zubiate et al., 2017; Cradden et al., 2017; H C Bloomfield et al., 2018) over Europe, consistent with the shifting path of extra-tropical cyclones travelling across the North Atlantic (Hurrell et al., 2003). During the NAO+, the anomalous warm and windy conditions over Northern Europe result in reduced demand and increased wind power generation (Ely et al.; Cradden et al., 2017; H C Bloomfield et al., 2018; Ravestein et al., 2018). In contrast, Southern Europe experiences anomalous low wind speeds that lead to reduced wind power generation (Jerez and Trigo, 2013; Zubiate et al., 2017). In general terms, the opposite is expected during NAO-. The above-average generation over Northern-Central Europe and below average generation in Southern Europe is consistent with the increased zonal flow conditions experienced during NAO+ (Grams et al., 2017).

The results presented by Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020) and Wiel, Hannah C Bloomfield, et al. (2019), exhibit a pattern characterized by warmer

and windier conditions in the north, over an extensive European land mass, and slightly less windier conditions in the south, and a weaker demand and residual load over central Europe.

(b) NAO-

Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020) and Wiel, Hannah C Bloomfield, et al. (2019) show that Northern Europe and Britain exhibit lower than normal winter temperatures consistent with a stronger than normal demand, whereas the areas of lower than normal wind speeds are restricted to Northern Europe. The demand and the residual load are near normal in most of Europe.

(c) AR

The AR pattern is related to anomalously northerly winds around the North Sea region and cold anomalies over central Europe. However, these temperature and wind anomalies are weak over land, and, as such, the pattern does not produce a significant large-scale response in either demand or residual load (Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez, 2020, Wiel, Hannah C Bloomfield, et al., 2019).

(d) SCAND

Studies confirm that the Scandinavian Blocking is associated with above-average demand and below-average wind generation in Central and Northern Europe (Grams et al., 2017; van der Wiel et al., 2019a). On the other hand, Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020) and Wiel, Hannah C Bloomfield, et al. (2019) exhibit a pattern associated with strengthened winds in the north and weakened winds in the North Sea, the Celtic Sea, and the Bay of Biscay; as well as anomalous warm temperatures in the north and anomalous cold temperatures in the south. The anomalies are weak over land and, analogous to the AR pattern, and there is no significant response in either the demand or the residual load.

2.3 K-means algorithm

The k-means is an unsupervised algorithm that iteratively finds the partition that minimizes the ratio of the variance within clusters to the variance between clusters centroids. These clusters are a collection of data points that are aggregated together according to certain similarities.

The target number k refers to the number of centroids, i.e., the center of the cluster. It has to be predefined. The first step of the algorithm consists of randomly selecting an initial set of centroids. Then, it performs iterative calculations to optimize the positions of the centroids until the centroids have stabilized or the maximum number of iterations has been achieved.

2.4 Joint probability distributions and dependence

When more than one random variable is defined, it is essential to distinguish between the joint probability distribution and the marginal probability distribution. The individual probability distribution of a random variable is referred to as its marginal probability distribution. The joint probability distribution is a probability distribution that gives the probability that each of the random variables falls in any particular range or discrete set of values specified for that variable.

The joint probability mass function of two discrete random variables X, Y is is:

$$p(X,Y)(x,y) = p(X = x \quad \text{and} \quad Y = y) \tag{2.1}$$

This is also called bivariate distribution, and, when generalizing to more than two variables, it is called multivariate distribution.

An event is a subset of the sample space. An independent event is an event that has no connection to the chances of another event to occur. In other words, the event noes not effect on the probability of another event occurring. Two random variables X and Y are independent if

$$p(X = x, Y = y) = p(X = x)p(Y = y) \tag{2.2}$$

It is necessary to highlight that, while the number of independent random events grows, the corresponding joint probability value decreases rapidly to zero, according to a negative exponential law.

When studying the joint behavior of random variables, it is useful to describe how they vary together by measuring the relationship between the variables. A common measure of the linear relationship between two random variables is the covariance. The covariance between the random variables X and Y is defined as

$$cov(X,Y) = [E(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y \tag{2.3}$$

Another measure of the linear relationship between random two random variables that is easier to interpret is the correlation. The correlation scales the covariance by the standard deviation (V) of each variable, which is defined as

$$\rho_{XY} = \frac{cov(X,Y)}{V(X)V(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{2.4}$$

2.5 Copulas

The goal of this thesis is to model the probabilities of co-occurrence of cold temperatures and weak winds. But estimating joint densities is not an easy task since only a few non-Gaussian families are defined, and non-parametric estimation is demanding. Nonetheless, density estimation in one variable is relatively easy, given the fact that many convenient families exist and that the non-parametric approach is efficient and accurate. The copulas framework for modeling multivariate distributions provides a flexible representation and separates univariates from the true nature of dependence.

In the field of probability theory and statistics, a copula function $C : [0,1]^n \to [0,1]$ is defined as a multivariate distribution

$$C(u_1, u_2, ..., u_n) = P(U_1 \leq u_1, U_2 \leq u_2, ...U_n \leq u_n) \tag{2.5}$$

such that marginalizing gives $U_i \sim Uniform(0,1)$. Copulas are useful because we can transform any arbitrary random variable into a uniform and back. The function that transforms uniforms to any other univariate distribution is the inverse of the cumulative density function (CDF). In order to do the opposite transformation, from an arbitrary distribution to the uniform(0, 1), we just apply the inverse of the inverse CDF, the CDF.

Let $X \sim F$ be a continuous random variable, then the distribution of $F(x) = P(X \leq x)$ is

$$\begin{aligned} P(F(X) \leq u) &= P(F^{-1}(F(X)) \leq F^{-1}(u)) \\ &= P(X \leq F^{-1}(u)) \\ &= F(F^{-1}(u)) = u \end{aligned} \tag{2.6}$$

Summarizing, the steps to create multivariate distributions from arbitrary marginal distributions are:

- Transform an arbitrary random variable, X to a uniform one $F_X(X)$, where F_X is the CDF of X.
- In the bivariate case, given a copula $C(U,V)$, U and V have uniform distributions. It is straightforward to extend the methodology to more variables.
- Given the random variables of interest, X and Y, a new distribution can be created as $C'(X,Y) = C(F_X(X), F_Y(Y))$.

2.5.1 Scope of the framework

Sklar's Theorem

An old mathematical result known as Sklar's theorem is the foundation of the concept of copula. It states that given an n-dimensional CDF, F, with marginals $F_1, F_2, ..., F_n$ there exists a copula function, C, such that

$$F_X(X_1, X_2, ..., X_n) = C_\theta(F_1(X_1,), F_1(X_2), ..., F_n(X_n)) \tag{2.7}$$

for all $X_i \in [-\inf, \inf]$ and $i = 1, ..., n$.

and if the marginals, F_i are continuous, the copula, C, is unique; otherwise, C is uniquely defined only on $Ran(F_1 \times Ran(F_2) \times ... \times Ran(F_n)$ where $Ran(F_i)$ denotes the range of the CDF.

Inversion of Sklar's theorem

In the opposite direction of Sklar's theorem, we have that given a copula, C, and univariate CDF's, $F_1, F_2, ..., F_n$. Then, F as defined in 2.7 is a multivariate CDF with marginals $F_1, F_2, ..., F_n$

2.5.2 Gaussian copula

As mentioned before, the advantage of modeling joint distributions with copulas is that, in practice, it is often easier to estimate the distribution of the marginals than to estimate the joint distribution. Copula theory ensures that, for every joint multivariate distribution, there exists a unique copula. In the case of the Gaussian copula function, finding its parameters is limited to finding the correlation matrix of the random variables we want to study.

A Gaussian copula is given by

$$C(u_1, u_2, ..., u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n)) \tag{2.8}$$

where Φ_Σ represents the CDF of a multivariate normal with covariance Σ and mean 0, and Φ^{-1} is the inverse CDF for the standard normal.

Given a multivariate distribution

$$F_X(X) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n) = \Phi_\Sigma(x_1, x_2, ..., x_n), \tag{2.9}$$

we can extract its Gaussian copula

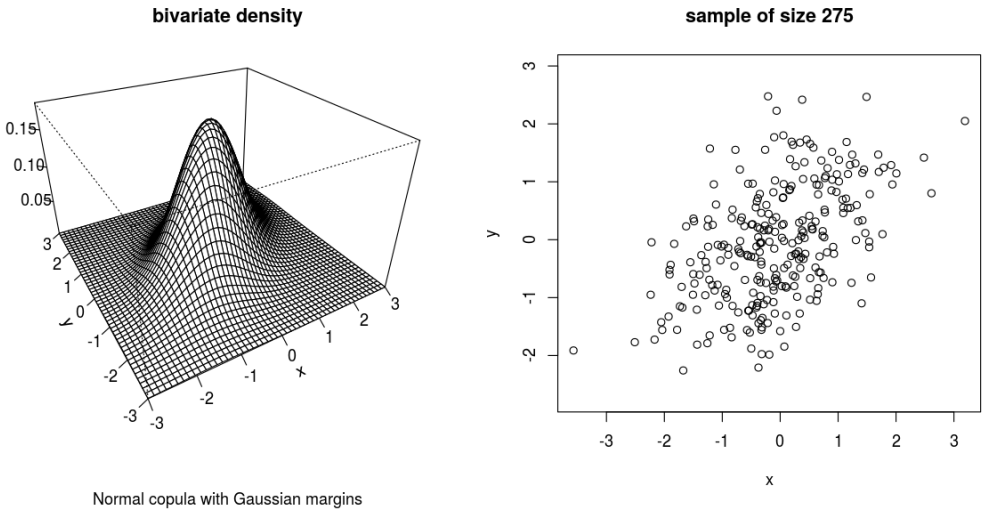

Figure 2.9: Density and scatterplot of a Bivariate Gaussian Distribution. The density of the joint distribution is obtained by joining a Gaussian Copula ($\rho = 0.5$) with two identical standard Gaussian univariate distributions. Sample size: $n = 275$. Gräler, n.d.

$$\begin{aligned} F_X(X) &= \Phi_\Sigma(F_1^{-1}(F_1(X)), F_2^{-1}(F_2(X)), ..., F_n^{-1}(F_n(X))) \\ &= \Phi_\Sigma(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_n^{-1}(u_n)) \\ &= \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n)) \\ &= C(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n)) \end{aligned} \tag{2.10}$$

and plug in any marginal into the copula function.

The inverse CDF transforms the uniforms to normal distributions, then, the multivariate normal's CDF squashes the uniform dimensions to be normally distributed. Thus, the Gaussian Copula is a distribution over the unit hypercube $[0,1]^n$ with uniform marginals.

The density plot of the Gaussian copula, as well as a scatterplot of the data generated using the tool <https://copulatheque.shinyapps.io/copulas/>, is provided in Figure 2.9. This figure was generate for the parameter $\rho = 0.5$, which is representative of the correlation values observed between low temperature and low wind speed events during the wintertime. The sample size is $n = 275$, approximately the number of events observed in a specific month during one of the WRs. For the same parameters, the strength of dependence and a scatterplot of the uniform variables are represented in Figure 2.10. It is clear that there is dependence. Further, the density is higher for (0,0) and (1,1), and lowest for (1,0) and (0,1), indicating a positive correlation.

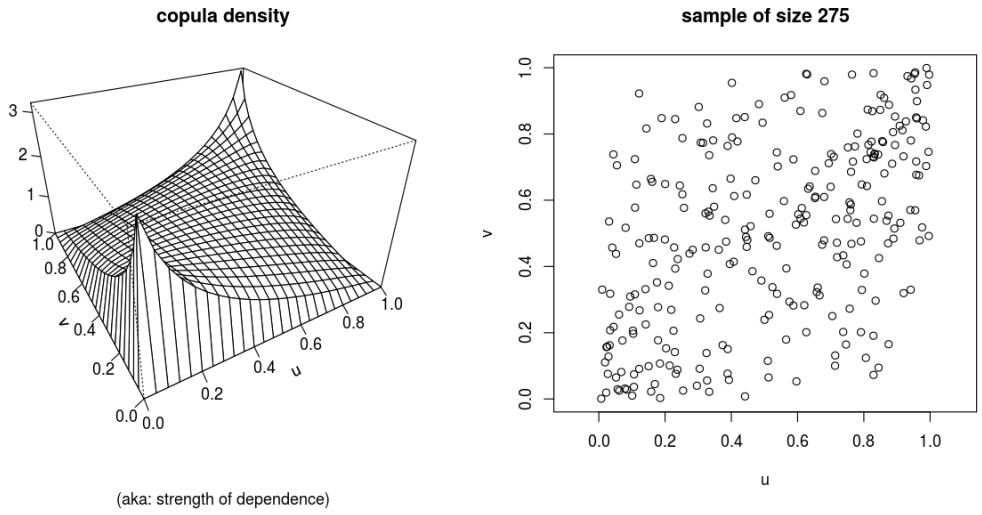

Figure 2.10: Density distribution and scatterplot of Gaussian copula ($\rho = 0.5$), with uniform marginal distributions. Sample size: $n = 275$. Gräler, n.d.

2.5.3 Semiparametric copula estimation

It has already been explained that the copula framework allows us to model the multivariate distributions by parameterizing the associations among the variables separately from their univariate marginal distributions. It is often the case that the marginal distributions do not belong to standard families. In such cases, it might be appropriate to use a semi-parametric strategy that involves representing the associations among variables with a simple parametric approach and estimating the marginals nonparametrically.

D. Hoff (2007) proposed an extended rank likelihood method of semiparametric inference for copula, which is a function of the association parameters only. It can be applied without any assumptions of the marginal distributions, making it appropriate for the joint analysis of continuous and ordinal discrete data. The package **cite** provides a tool for estimation and inference for the Gaussian copula parameters via a Markov chain Monte Carlo algorithm based on Gibbs sampling.

Let y_1 and y_2 be two random variables with continuous CDFs F_1 and F_2. The transformed variables u_1 and u_2 both have uniform marginal distributions. A semiparametric copula model includes a parametric model for the joint distribution of u_1 and u_2, but lacks any parametric restrictions on F_1 and F_2. Details about how the likelihood function that depends on association parameter and not on the unknown marginal distribution is derived can be found in D. Hoff 2007, more specific, in section 3 of this paper.

2.6 Linear regression

As explained before, modeling the events' marginal distributions is needed to model the joint distributions with the copula function. In this work, a linear regression model is proposed to model the univariate probabilities of low temperature events and low wind speed events.

Linear regression is a useful tool for predicting a quantitative response Y on the basis of a set of p predictors $X_1, X_2, ..., X_p$, assuming a linear relationship. The linear regression model takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon, \tag{2.11}$$

where the intercept, β_0, and the slope terms, $\beta_1, ..., \beta_p$ are unknown constants. The parameters are estimated using the least square approach.

2.6.1 Logistic regression

Logistic regression is a widely used linear model for binary classification. To describe this model, it is convenient to introduce the logit function defined as the natural logarithm of the odds ratio (log-odds). Let Y be the binary outcome variable indicating failure or success with $0, 1$. Then p stands for the probability of a positive event, i.e., $p = P(Y = 1)$, and the quotient $p/(1-p)$ is the odds ratio: the odds in favour of a particular event. The mathematical expression of the logit function is:

$$logit(p) = log(\frac{p}{1-p}), \tag{2.12}$$

The logit function takes probability values in the range [0, 1] and transforms them into values over the entire real-number range. The reason why this transformation is applied is that it is usually difficult to model a variable that has a restricted range.

When a logistic regression model is applied, it is assumed that the logit transformation of the outcome variable has a linear relationship with the predictor variables. Let Let $x_1, x_2, ..., x_k$ be a set of predictor variables. Then $\beta_0, \beta_1, ..., \beta_k$ are the parameters estimated via the maximum likelihood method when performing a logistic regression of Y on $x_1, x_2, ..., x_k$:

$$logit(p) = log(\frac{p}{1-p}) = \beta_0, \beta_1 x_1, ..., \beta_k x_k. \tag{2.13}$$

We are usually interested in predicting the probability that a particular sample belongs to a particular class. First, we exponentiate and take the multiplicative inverse of 2.13:

$$\frac{1-p}{p} = \frac{1}{\beta_0, \beta_1 x_1, ..., \beta_k x_k}. \tag{2.14}$$

Then, partial out the fraction on the left-hand side of the equation and add one to both sides, obtaining:

$$\frac{1}{p} = 1 + \frac{1}{e^{\beta_0, \beta_1 x_1, ..., \beta_k x_k}}. \tag{2.15}$$

Finally, we change 1 to a common denominator and take the multiplicative inverse to obtain the formula of the probability $P(Y = 1)$:

$$p = \frac{e^{\beta_0, \beta_1 x_1, ..., \beta_k x_k}}{1 + e^{\beta_0, \beta_1 x_1, ..., \beta_k x_k}} = \frac{1}{1 + e^{-(\beta_0, \beta_1 x_1, ..., \beta_k x_k)}}, \tag{2.16}$$

denominated as logistic sigmoid function, or simply sigmoid function due to its characteristic S-shape.

2.7 Brier Score

Forecast performance is often assessed with a scalar summary measure, which is practical but gives an incomplete picture (Wilks, 2006). Numerous scalar measures for probabilistic forecasts exist but the most common is the Brier Score (BS) (Brier, 1950).

It is assumed that the events only can occur in one of r classes on each of the n occasions, and the forecast probabilities of each classes are $f_{i1}, f_{i2}, ..., f_{ir}$ in one of the occasions, i. The r classes are chosen to be mutually exclusive and exhaustive, so that

$$\sum_{j=1}^{r} f_{ij}, \quad i = 1, 2, 3, ..., n \tag{2.17}$$

The definition of the verification score P proposed by (Brier, 1950) is

$$P = \frac{1}{n} \sum_{j=1}^{r} \sum_{i=1}^{n} (f_{ij} - E_{ij})^2 \tag{2.18}$$

where E_{ij}, takes the value 1 if the event occurred in class j and 0 otherwise. In the case of probabilistic forecasts of dichotomous events, the score averages the squared differences between the pairs of forecast probabilities and the binary observations. Thus the equation for the verification score is

$$BS = \frac{1}{n} \sum_{i=1}^{n} (f_i - E_i)^2 \tag{2.19}$$

which is analogous to the equation for the mean squared error, and negatively oriented, with perfect forecasts exhibiting $BS = 0$. Less accurate forecasts exhibit higher scores, but since individual forecasts and observations are both bounded by zero and one, the range of possible values for the BS is $0 \leq BS \leq 1$. The expression of the BS in (2.19) is extensively used, but it differs from the original score introduced by (2.18), in that it averages only the squared differences of one of the two classes.

It is important to highlight that the scores favour the use of some skill in forecasting an average departure from climatological probabilities, but in the complete absence of any forecasting skill, predicting climatological probabilities will lead to a better score than categorically forecasting the most frequent class on every occasion. The BS also encourages to state unbiased estimates of the probability of each event when the forecast cannot be exactly right.

The Brier Skill Score (BSS) is often used and, since $BS_{perf} = 0$, it takes the form

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \tag{2.20}$$

The reference model is usually the climatology. Negative values mean that the forecast is less accurate than the reference forecast; when the forecast presents no skill compared to the reference $BSS = 0$; and a perfect skill compared to the reference forecast reflects in a skill score equal to 1.

2.8 Permutation test

Permutation tests are a group of nonparametric test procedures. Here, they are employed to test the null hypothesis that two independent groups come from the same distribution. They can be performed with virtually any measure of location or scale.

Permutation tests depend on the principle of exchangeability, which implies that all the data were drawn from the same distribution under the null hypothesis (Wilks, 2006). In other words, under the null hypothesis (treatment = control), any permutations are equally likely. For a two-sided test, the alternative hypothesis is defined so that the two samples are different. If only the extreme chance results in one direction count toward the p-value, a one-tail test can be performed.

No sampling distribution is assumed; it is instead built by resampling the observed data without replacement. Since the samples are drawn without replacement, each of the individual n observations is only represented once in one of the artificial samples of size $n1$ and $n2$, and the data labels are randomly permuted for each resample. The test statistic is computed for each

pair of these samples and is compared to the resulting distribution outcomes that form the null distribution.

Let X_1 be a sample of size n_1 from an unknown distribution F, the treatment group, and let X_2 be an independent sample of size n_2 from an unknown population distribution G, the control group. Suppose that we want to use the data to test the null hypothesis that the two population distributions are identical.

The algorithm can be implemented as follows:

- Compute the ground truth absolute difference between the samples mean (or using another metric) $d = |\bar{X}_1 - \bar{X}_2|$.
- Pool the variables into one single distribution.
- Consider any permutation of the pooled data. Compute the sample mean of the first observations and that of the sample mean using the remaining observations. Then, compute the difference between these sample means.
- Repeat the previous step for all possible permutations of the data yielding, say, L differences: $\hat{\delta}_1, \hat{\delta}_2, ..., \hat{\delta}_L$. Given that computing all the permutations is expensive, in practice, only p pairs of random samples are considered.
- The significance value is the proportion of permuted differences higher than the ground truth difference calculated in the first step. Here, the p-value is defined as the probability, given the null hypothesis is true, that we obtain results that are at least as extreme as the results we observed (the ground truth).

3 Data and Methods

3.1 Data

3.1.1 Reanalysis

Climate reanalysis combines past observations with models to generate time series of climate variables. In this study, ERA5 reanalysis (Hersbach et al., 2019) was chosen to represent observed historical meteorological conditions, spanning the period 1980-2017. This is the latest climate reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and replaces the ERA-Interim reanalysis, which stopped being produced in 2019. ERA5 is based on 4D-Var data assimilation using Cycle 41r2 of the Integrated Forecasting System IFS. It provides hourly estimates of a large number of atmospheric and oceanic variables together with uncertainty parameters. The data covers the Earth and is available on 37 pressure levels and single levels, on a regular latitude-longitude grid of $0.25^\circ \times 0.25^\circ$ resolution (Hersbach et al., 2019). All the Python scripts used to download the data files in NetCDF format are available in the GitHub repository.

Temperature is well established as the main weather driver of electricity demand. Very cold or warm temperatures increase the demand due to heating and cooling, respectively (Taylor and Buizza, 2003). Here, the minimum daily temperature was calculated from hourly air temperature at 2 meters above the surface of land, sea, or inland waters. This variable is computed by interpolating between the lowest model level and the Earth's surface, taking into account the atmospheric conditions. Climatology values (1980-2018) of minimum daily temperature across the extended European domain during the boreal winter season (November to March) are presented in Figure 3.1.

Wind speed was calculated from the eastward (u) and northward (v) components of the wind at 10 meters at an hourly frequency, and then aggregated to daily maximum values. The ECMWF warns about taking care when comparing these parameters with observations, which can vary on small scales and time scales, and are affected by the local terrain, vegetation and buildings, represented only on average in the ECMWF IFS. Climatology values (1980-2018) of maximum daily wind speed during the boreal winter season from the ERA5 dataset are presented in Figure 3.2.

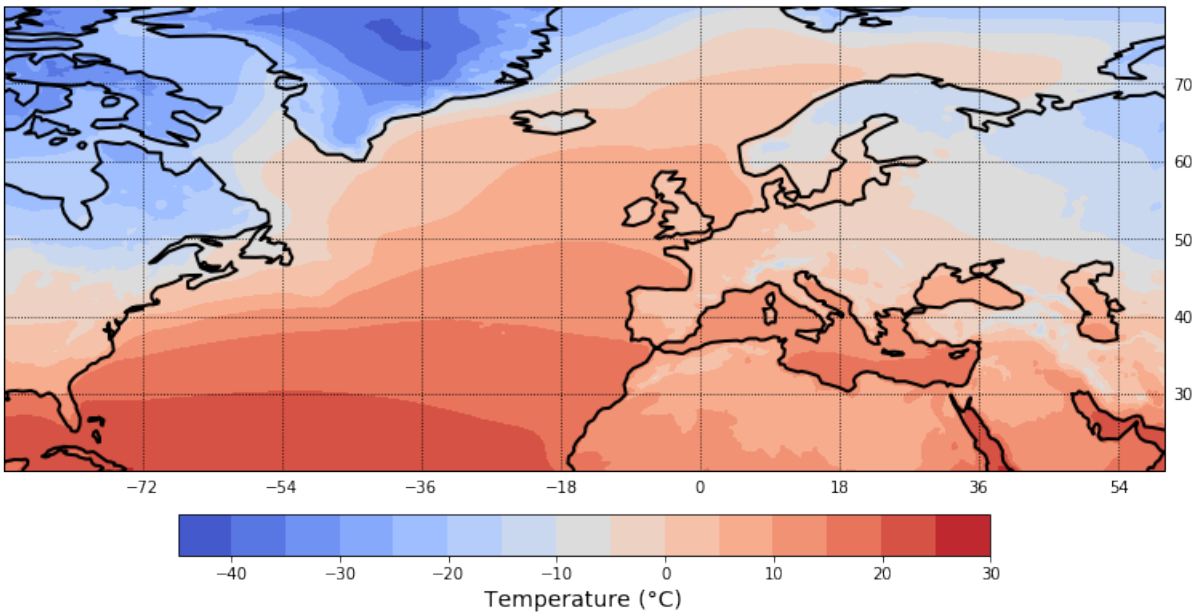

Figure 3.1: ERA5 minimum daily temperature climatology at 2 meters height. Figure based on ERA5data (NDJFM, 1979–2017).

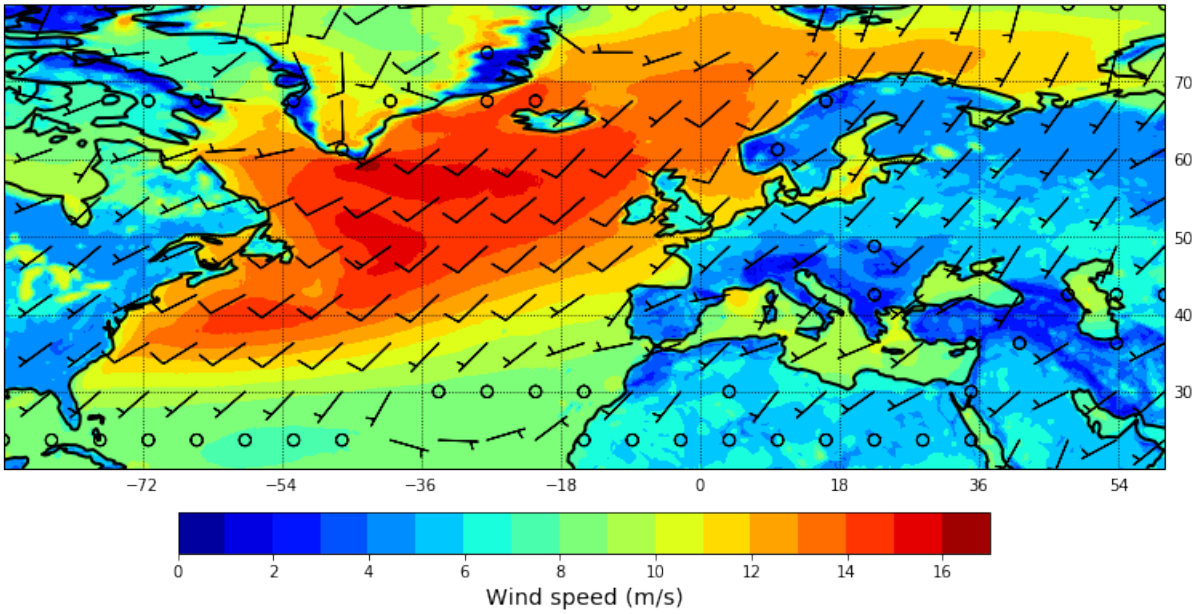

Figure 3.2: ERA5 maximum daily wind climatology at 10 meters height. Colours show maximum daily wind speeds. Figure based on ERA5data (NDJFM, 1979–2017).

3.1.2 Weather Regimes

WR patterns were computed following the classification method proposed by Cassou. Z500 data from ERA5 is used, constrained to the Euro-Atlantic region (27°-81°N, 85.5°W-45°E), during the winter season (November to March) in the period 1980-2017. The WRs were computed

according to Wiel, Hannah C Bloomfield, et al. (2019 and Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020. The calculation method consists of two steps. In the first stage, a cosine weight as a function of latitude is applied to the, data and the first fourteen EOFs patterns are computed. The associated Principle Component (PCs) time series were used as coordinates of a reduced phase space. Then, the PCs were clustered into four groups with the K-means algorithm, choosing 30 random starts and a maximum of 100 iterations. Every daily map was assigned to a centroid (fig. 3.3) based on its closest distance (Euclidian in our case). The number of regimes k=4 was chosen as it corresponds to the more robust regime partition during winter months (Michelangeli, Robert Vautard, and Legras, 1995). The Z500 anomalies for each cluster are exhibited in Figure 3.3, and the proportion of days in each cluster is approximately:

- AR: 22%
- SCAND: 28%
- NAO-: 19%
- NAO+: 32%

.

3.2 Methodology

This thesis's primary goal is to design a methodology that allows for the computation of the joint probabilities of occurrence of low temperature and low wind speed events, in order to identify meteorological conditions that could eventually lead to energy shortfall. The results were achieved by defining the low temperature events as daily minimum temperatures below the 10^{th} percentile derived from the seasonal (Nov-March) distribution. Analogously, low wind speed events are defined by the daily maximum wind speeds below the seasonal 10^{th} percentile. This is the simplest way of defining events, but other formulations can be adopted in the current context, given that the methodology is independent of the arbitrary thresholds defined in this thesis. A comparison of the modeled marginal probabilities for different thresholds is provided in Appendix 8. Daily values are computed from hourly data.

Research in this field has focused on average meteorological conditions and departures from seasonal means (e.g. Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez, 2020;Cortesi et al., 2019; Wiel, Hannah C Bloomfield, et al., 2019. In contrast to that approach,

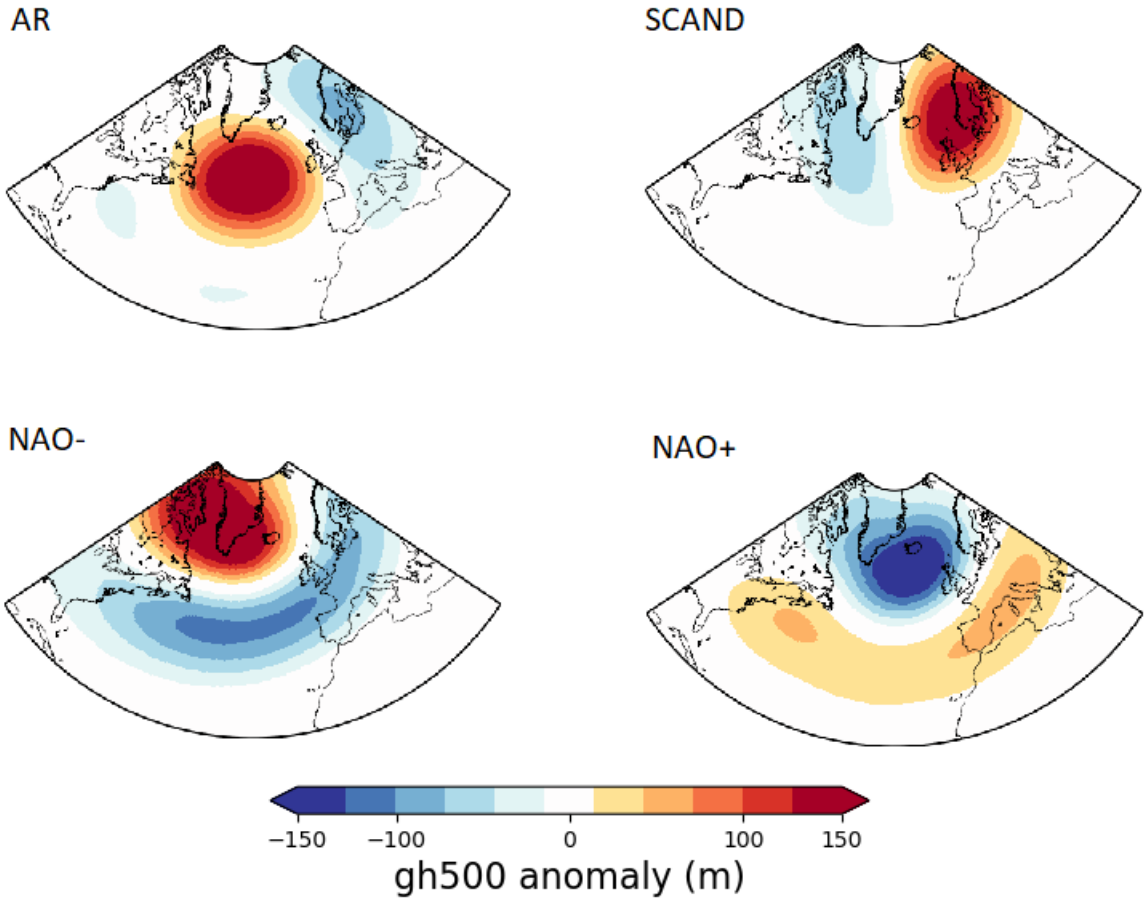

Figure 3.3: Four regimes of atmospheric circulation in the North Atlantic-European domain,AR, SCAND, NAO-, NAO+. Colours show the Z500 anomaly (m). Area of study: 27N-81N, 95.5W-45E. Figure based on ERA5 data (NDJFM, 1979–2017).

the methodology described in this thesis benefits from the high resolution of the ERA5 datasets to predict the joint probabilities. For the sake of comparison, the average meteorological surface impact for each regime was determined through composite analysis, that is, the mean over all days classified in the regime. Anomalies were computed by subtracting the daily climatology.

Empirical marginal and joint distributions of daily temperature and wind speed events were calculated for the entire season and for each month separately. The importance of the WRs defined by Cassou, 2008 in the prediction of compound events is manifested in the analysis of the marginal probabilities of occurrence of low temperature and low wind speed events. Together with the correlations between these two variables, modeled marginal probabilities are embedded in Gaussian copula functions employed in the generation of the joint probabilities on a grid level.

A logistic regression model was adopted to estimate the marginal probabilities of low daily wind speeds or low temperatures at each location, s. For each variable, the model shows the interaction effect of regressing the dependent variable on a set of five dummy variables for the months in the season November-March, $M_{i,t,s}$, $i \in \{Nov, Dec, Jan, Feb, Mar\}$ and a set of four dummy variables represents the dependence on the four weather regimes, $R_{j,t,s}$, $j \in \{AR, SCAND, NAO-, NAO+\}$. At any time t, in this case a day, one of the seasonal dummies $M_{i,t,s}$ will equal 1, while all the others will equal 0. No intercept term is added to the model to avoid collinearity. The model can be expressed as

$$logit(P(\hat{Y}_{t,s} = 1)) = \sum_{i,j} \beta_{i,j} M_{i,t,s} R_{j,t,s}, \tag{3.1}$$

where $\beta_{i,j}$ are the regression coefficients, and P is the probability that $\hat{Y}_{t,s}$, the variable of interest (either the minimum daily temperature at two meters or the maximum daily wind speed at 10 meters), is below an arbitrary threshold. In other terms, there are two possible categories defined by the threshold, in this case, the 10^{th} percentile.

$$\hat{Y}_{t,s} = \begin{cases} 1, & \text{if the observation is below the threshold} \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

The reference model to which the results of the logistic regression model were compared to is a simple climatology model consisting of the main effect of regressing $P(\hat{Y}_{t,s} = 1)$ on $M_{i,t,s}$ as defined above, given by

$$logit(P(\hat{Y}_{t,s} = 1)) = \sum_{i} \gamma_i M_{i,t,s}, \tag{3.3}$$

where $\gamma_{i,j}$ are the regression coefficients.

To illustrate the importance of using in the prediction of low temperature and low wind speed events, a logistic regression model with the WRs as the only predictors was run for the winter season.

The univariate probabilities are estimated using cross-validation, using the leave one group out method, being the group defined by one year of data. This means that, for each year, the coefficients of the logistic regressions were estimated using the remaining data.

The marginal probabilities from the main model were compared to the climatology model in terms of the BSS. The BSS was computed with one of many possible samples from the population, therefore the sampling uncertainty affecting this metric needs to be taken into account. The

significance of the results was assessed with a permutation test, using the BS difference between the main model and the climatology model at each location as a metric, and running $p = 1000$ permutations. All the tests were performed at a 10% significance level. The methodology for computing the permutation test at each location can be summarized as follows:

- Subset the samples by selecting the predictions that are classified in the regime that is analyzed. For each model, the treatment group is defined by the predictions and the control group is defined by the observed values.
- Compute the ground truth as $gT = BS_{main_model} - BS_{climat_model}$.
- For each model, pool the treatment and the control variables into one single distribution.
- Shuffle the data $p = 1000$ times, and store the permuted difference of BS in a new variable.
- Compute the significance values as the proportion of permuted differences higher than the ground truth.

The joint probabilities were estimated at each location using a Gaussian copula function that takes as an input the marginal probabilities and the covariance matrix for the two variables. Correlations are estimated using the function mcmc from the R package 'sbgcop' (D. Hoff, 2007). This function produces MCMC samples from the posterior distribution of a correlation matrix, using a scaled inverse-Wishart prior distribution and an extended rank likelihood. The importance of consireing the associations between the events when modeling the joint distributions is assessed by computing the BSS using a model with no correlation as a reference and performing a permutation test at the 10% significance level.

The code for reproducing the results is available in the GitHub repository paulina-t/msc-thesis. The experiments are conducted on one of the servers of the Department of Geosciences from the University of Oslo, Wessel. Its Characteristics are provided in the following table. The Linux version installed is 3.10.0-1127.18.2.el7.

4 Results

4.1 Mean meteorological conditions

Composites were computed for minimum daily temperature and maximum daily wind speed anomalies. Figures 4.1 and 4.2 show the typical surface imprint of the four WRs on temperature and wind speed anomalies respectively. The mean meteorological conditions illustrated in both figures are in good agreement with previous findings in the literature (e.g. Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez, 2020; Wiel, Hannah C Bloomfield, et al., 2019).

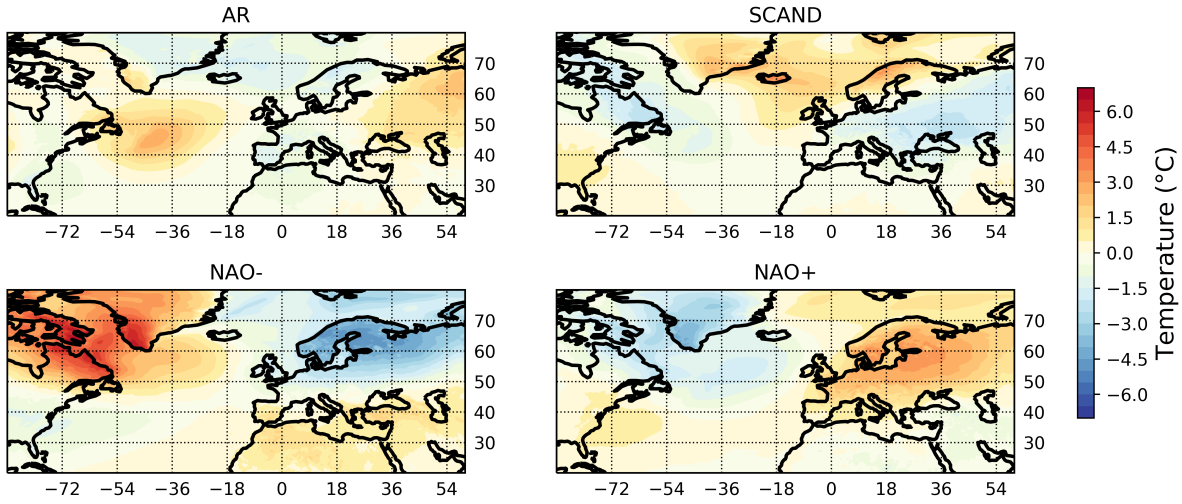

Figure 4.1: Mean meteorological surface impacts of the four WRs. Colours show maximum daily temperature anomalies (°C). Area of study: 20N-80N, 90W-60E. Figure based on ERA5 data (NDJFM, 1979–2017).

As expected, the regimes with the strongest impact are NAO- and NAO+. The response of these regimes in both the temperature and the wind speed fields is symmetric. Anomalous low surface temperatures over Europe are usual during blocking conditions, in this context, during NAO- and SCAND. The impacts of the NAO on surface temperature (Figure 4.1) consists of a zonal dipole with centers over Northern Europe and the Labrador Sea. During NAO-, temperature anomalies are, on average, cold across Northern Europe, Scandinavia, and Northwestern Russia, but warm across the Mediterranean countries. This results from above-average geopotential heights over Iceland, allowing cold air to drain from high latitudes, and a lower than

normal pressure system over the Azores high. These effects combined, decrease the pressure gradient across the North Atlantic. Meanwhile, the opposite conditions yield for NAO+, and warm anomalies cover the entire continent. A north-south gradient of mean temperature anomalies can be observed during SCAND conditions, but less prominent and with the opposite sign of the gradient associated with NAO-. The anticyclonic circulation associated with SCAND imposes warm conditions over Scandinavia and Britain, extending over the Norwegian Sea to Greenland, as a cold polar continental air mass develops across the rest of Europe. The lowest climatological departures of minimum daily temperatures are observed during AR days. This pattern consists of two centers of warm anomalies located over the North Atlantic and Northwestern Russia, and cold anomalies over the European continent spanning the Norwegian Sea and Greenland.

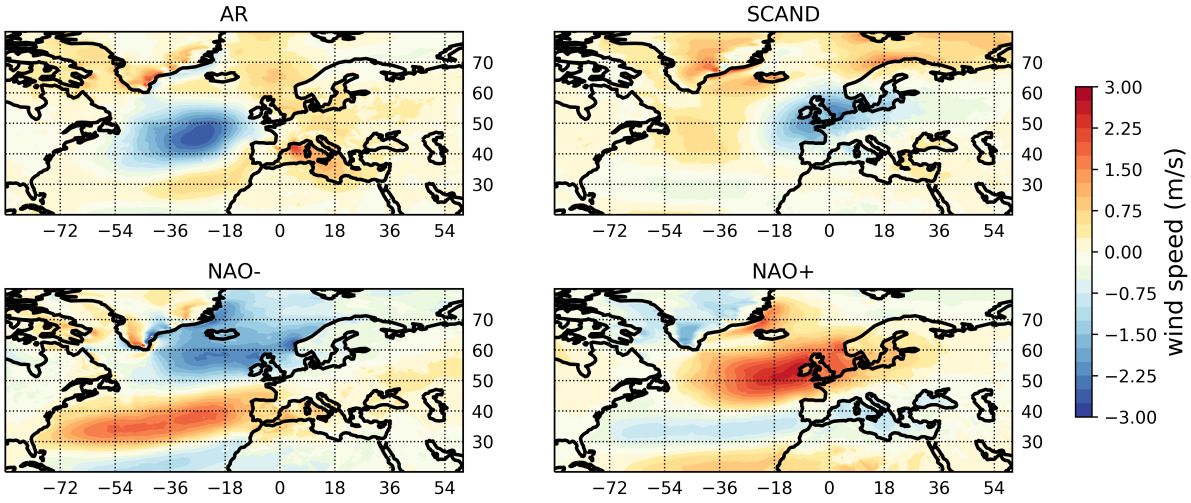

Figure 4.2: Mean meteorological surface impacts of the four WRs. Colours show minimum daily wind speed anomalies (m/s). Area of study: 20N-80N, 90W-60E. Figure based on ERA5 data (NDJFM, 1979–2017).

Low wind speed conditions over Europe occur on average during blocking situations when typically a high pressure system dominates the circulation over Europe, i.e., NAO- and SCAND (see Figure 4.2). During NAO, a meridional wave train extends across the Atlantic and Western Europe with an opposite sign for the positive and negative phases. During NAO+, countries in Northern Europe are on average affected by high anomalous wind speeds, whereas Mediterranean countries experience anomalous low wind conditions; the opposite yields during NAO-. The SCAND is characterized by anticyclonic anomalies resulting in anomalous weak winds centered over Britain and affecting most European countries. Similarly, the impacts of the AR

on the wind speed anomaly field consists of a monopole of weak winds but, in this case, with a center located in the North Atlantic with strong wind anomalies spanning the European continent.

4.2 Proportion of meteorological events by WR

Meteorological events are defined on each grid box as the days when the variable in concern is below the 10^{th} percentile. The threshold is computed at each location based on all the daily values in the extended winter season. For the rest of the analysis, a smaller region covering the European continent is selected. It extends from 35N to 72N and from 11W to 40E.

4.2.1 Temperature events

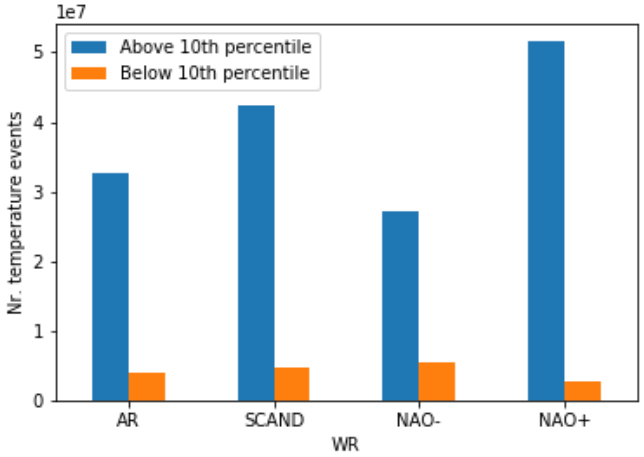

Figure 4.3: Total number of temperature observations by regime. Blue columns represent observations above the 10th percentile computed for the winter season (NDJFM) on a grid level, and orange columns represent observations below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

The sum over all locations of the number of days classified in each regime is depicted in Figure 4.3, using orange bars for the number of temperature events and blue bars for the remaining days. Complementing this figure, Table 4.1 provides the percentage of days represented in each column with respect to the total number of days classified on each cluster, together with the ratio ($\frac{\%events}{\%notevents}$), and the total number of days in each regime. Given that the events were defined as the number of days with temperatures below the 10^{th} percentile, the sum of days represented by the four orange bars is the 10% of the entire winter season. Consequently, the values in the second column of Table 4.1 sum 10, and the values in the first column sum 90.

It can be observed that the NAO+ is the dominant pattern in Europe, accounting for almost

Table 4.1: Proportions of events in each WR computed over the entire European domain. The first column, $\%t > q$, represents the percentage of the total number of days with temperatures above the threshold over the entire grid, computed with respect to the total number of days in each cluster. Analogously, the second column, $\%t < q$, represents the percentage of the total number of events in Europe. The ratio between these two columns is written in the third column. The last column is the sum of the values in the first two columns and represents the percentage of the number of days classified in each WR. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

WR	%(t > q)	%(t < q)	ratio	sum
AR	19.11	2.40	0.13	21.52
SCAND	24.75	2.78	0.11	27.54
NAO-	15.91	3.22	0.20	19.14
NAO+	30.21	1.60	0.05	31.81

1/3 of the days. The SCAND pattern is also frequent, as 28% of the days are classified in this cluster. Although NAO+ is the most frequent pattern, the ratio of cold events is the lowest among the clusters, 5:100. This supports the notion that the NAO+ is associated with above-normal temperatures in Europe (see Figure 4.1. On the contrary, NAO- is the less observed pattern but presents the highest ratio of events, 2:10, which is consistent with lower than normal temperatures observed during NAO-. The SCAND is the second most frequent regime, but the ratio of events of the AR is lower than for the NAO-. This is also in line with Figure 4.1 that shows higher temperature anomalies for both phases of the NAO compared tho the other clusters.

4.2.2 Wind speed events

Figure 4.4 and Table 4.2 are constructed as Figure 4.3 and Table 4.1 but for weak wind events. The last column is independent of the variable and is, therefore, the same. The number of events is, however, different. Furthermore, there is less dispersion among the WRs in the case of wind speed events. The clusters with the highest ratios are, as expected, the blocking patterns, SCAND and NAO-, with proportions 14:100 and 12:100 respectively. The SCAND is also the

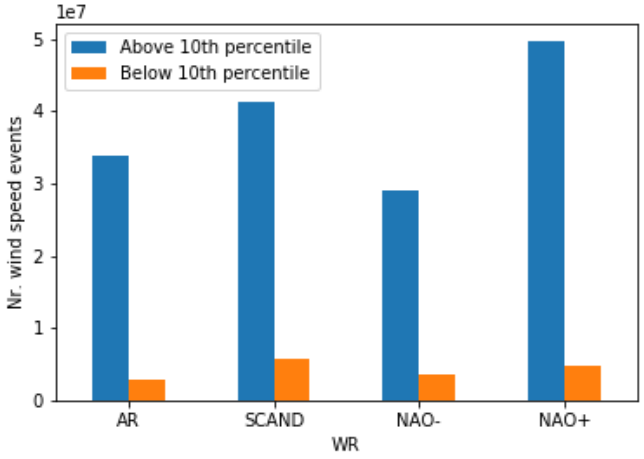

Figure 4.4: As Figure 4.1 but for the wind speed variable. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table 4.2: As Tanble 4.1 but for wind speed events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

WR	%(ws > q)	%(ws < q)	ratio	sum
AR	19.80	1.71	0.09	21.52
SCAND	24.11	3.43	0.14	27.54
NAO-	17.02	2.12	0.12	19.14
NAO+	29.07	2.74	0.09	31.81

pattern with the highest frequency of weak wind events, coherent with a high pressure system covering a large fraction of Europe. The AR and the NAO+ have both ratios 9:100.

4.2.3 Joint temperature and wind speed events

Figure 4.5 and Table 4.3 are obtained by combining the results presented above. The total number of joint events illustrated in Figure 4.5 is highest for the blocking patterns, and lowest for the NAO+. Although the total number of events associated with NAO- and AR is similar, the ratios are considerably different. In the case of NAO- events, the ratio is 24:1000, whereas in the case of SCAND, it is 15:1000. In addition, the ratio of joint events for the NAO+ is an order of magnitude smaller than for the other events. The reason why the difference in the total

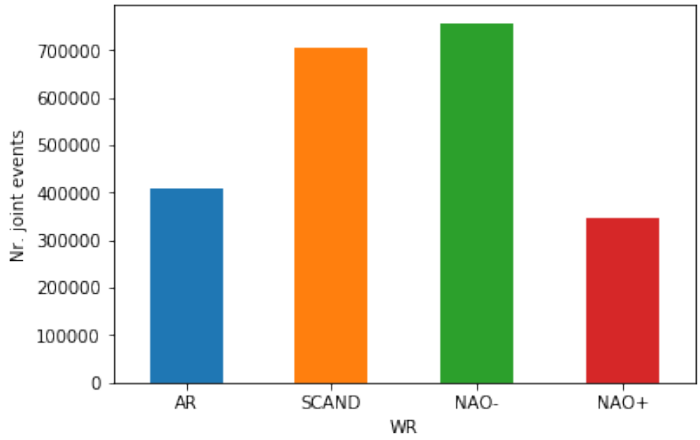

Figure 4.5: Total number of joint temperature and wind speed events by regime over the European domain. The threshold is computed on a grid level for each variable separately as the 10^{th} percentile considering all the days in the winter season (NDJFM). Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table 4.3: As Table 4.1 but for joint temperature and wind speed events. The threshold for defining cold events and weak wind events is the 10^{th} percentile of each variable computed for all the days in the winter season (NDJFM). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

wr	% not joint events	% joint events	ratio	sum
0.0	21.28	0.24	0.01	21.52
1.0	27.12	0.41	0.02	27.54
2.0	18.70	0.44	0.02	19.14
3.0	31.61	0.20	0.01	31.81

number of events between the NAO+ and the rest of the clusters is not as big as the ratio is that more days are classified as NAO+.

The results presented above confirm the hypothesis that blocking patterns lead to a higher number of joint low temperature and low wind speed events, due to anomalous high pressures affecting the circulation across Europe. , characterized by the opposite conditions, leads to the lowest frequency of joint events.

4.3 Empirical distributions

The results exhibited in the previous section are representative of the entire domain. The spatial distribution is studied carefully in this section. Empirical marginal and joint probabilities were computed as the observed frequencies of events relative to the number of days in each WR. The results are provided for the winter season extending from November to March, and for each month separately showing the seasonal behaviour. The following notation is adopted to describe the results: p denotes the marginal probability of occurrence, and the subindices X, N, and μ refer to the maximum, minimum, and average probabilities over all locations. The month and/or regime is in parenthesis.

4.3.1 Marginal probabilities

Seasonality of marginal probabilities

The seasonality of the marginal probabilities of low temperature and low wind speed events was studied at a monthly level. The results are illustrated in the present subsection.

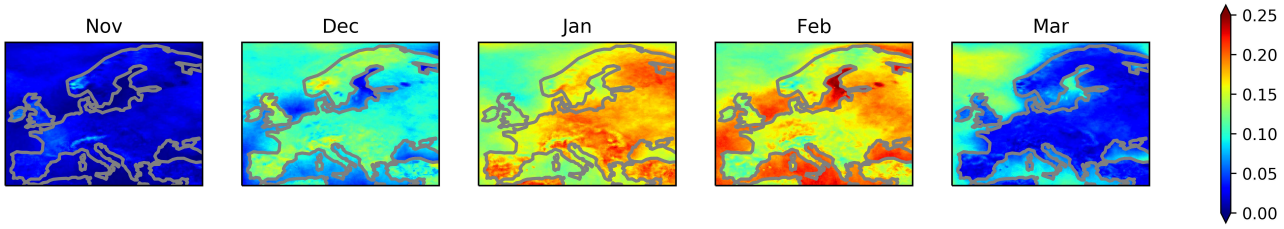

Figure 4.6: Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Temperature The temperature variable exhibits the expected pattern, with the warmest temperatures in the transition months, November and March. The coldest surface temperatures over land occur more frequently in January ($p_{X_{(Jan)}} = 0.2328$, $p_{\mu_{(Nov)}} = 0.1597$), whereas the coldest temperatures over the sea are observed in February ($p_{X_{(Feb)}} = 0.2574$, $p_{\mu_{(Feb)}} = 0.1654$). The response to heating over land is faster than over water bodies due to the water's higher heat capacity. A monthly lag between the distinct surfaces can be observed in Figure 4.6. November is the warmest month, the maximum probability of low temperatures is $p_{X_{(Nov)}} = 0.1000$ and the averages over all grid points are $p_{\mu_{(Nov)}} = 0.0188$. The maximum registered in March is

$p_{X_{(Mar)}} = 0.1630$, whereas the average is $p_{\mu_{(Mar)}} = 0.0634$.

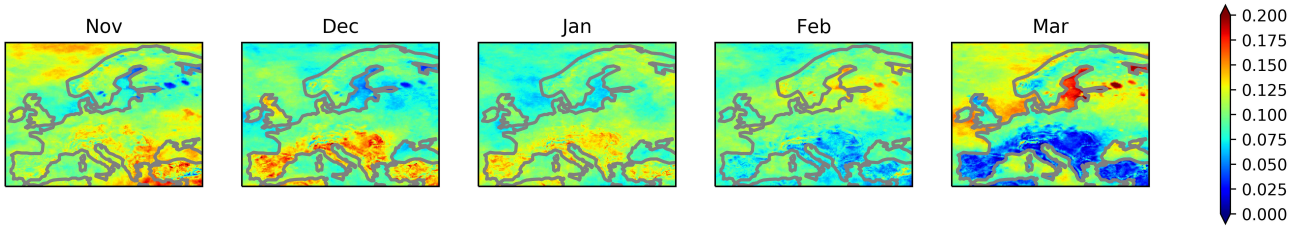

Figure 4.7: Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Wind speed As regards the wind speed variable, the seasonality is more complex than the described for the temperature variable. In Southern Europe, observed probabilities of low wind speeds are high from November to January, being the highest probabilities registered in December ($p_{X_{(Dec)}} = 0.2084$, $p_{\mu_{(Dec)}} = 0.0978$). The latitudinal gradient shifts sign in February, and the lowest values in the south are observed in March. On the other hand, in March and November, the warmest months of the season, probabilities are high over the ocean in high latitudes. The maximum values observed in March are $p_{X_{(March)}} = 0.2180$ and $p_{\mu_{(march)}} = 0.0990$. March is also the month with the biggest contrasts between ocean and land. Furthermore, the probability values are lower than for temperature events.

Marginal probabilities computed by regime

The importance considering the WRs in the study of marginal probabilities of low wind speed and low temperatures is evident in Figures 4.6 and 4.7. Clear patterns associated with each of the WRs are depicted in these figures.

Temperature As figure 4.8 indicates, the highest probabilities of occurrence of low temperatures are associated with the NAO- in Northern Europe and Scandinavia, in particular, on the west coast of Norway, Great Britain, and Ireland. The maximum over all grid boxes is $p_{X_{(NAO-)}} = 0.3100$, which is higher than any of the monthly maximum values in Figure 4.6. The pattern extends from southwest to northeast, with the lowest probabilities in the Mediterranean. During AR conditions, a maximum is visible over the North Atlantic ($P_{X_{(AR)}} = 0.2243$), and relatively high probabilities are observed in the Mediterranean. However, most continental regions

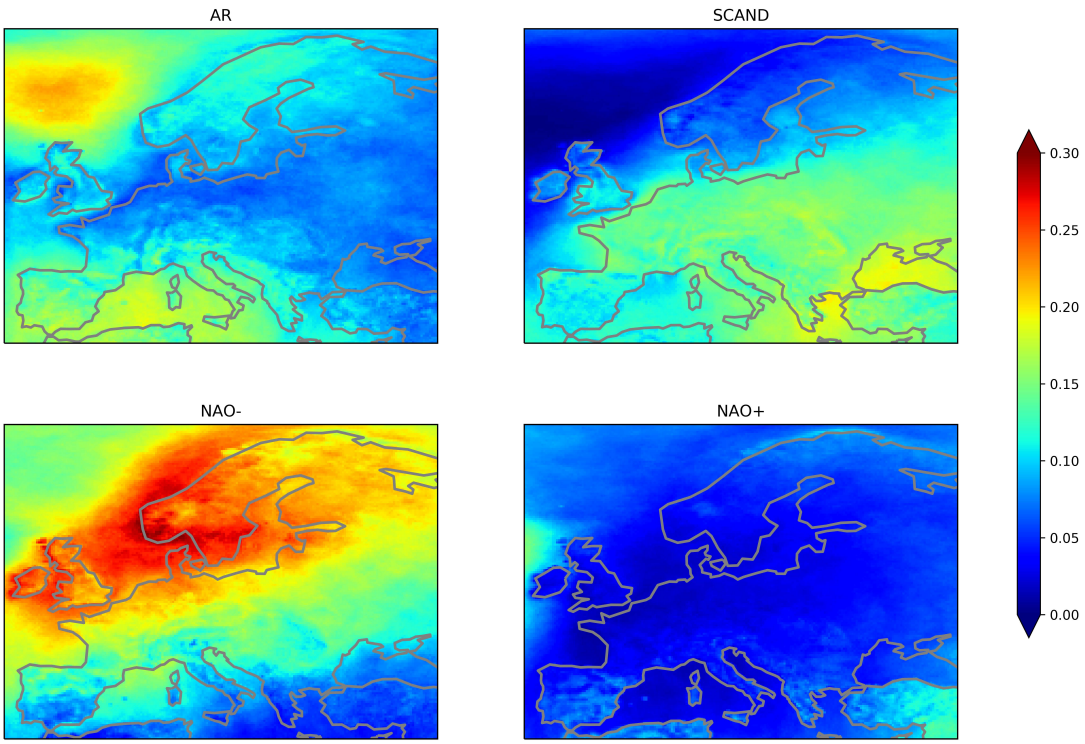

Figure 4.8: Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} *percentile for each WR. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

show no relevant signal. The SCAND pattern illustrates a NW-SE gradient, with a maximum in the Aegean and Black seas and no observed events in the North Atlantic ($p_{N_{(SCAND)}} = 0$). Cold conditions are expected to occur with less frequency during NAO+ compared with the other clusters. The overall spatial maximum and minimum for this regime are $p_{X_{(NAO+)}} = 0.1556$ and $p_{N_{NAO+}} = 0.0062$ respectively. In contrast, NAO+ is the pattern with the lowest observed frequency ($p_{\mu_{(NAO+)}} = 0.0502$).

Overall, NAO- is the pattern with the highest average of observed low temperature events ($p_{\mu_{(NAO-)}} = 0.1685$), followed by the AR. Meanwhile, the spatial distribution is considerably different and should be taken into account in the design of energy systems. Days classified as NAO- are characterized by the highest probabilities in Northern Europe and Scandinavia. The SCAND shows a similar pattern, but weaker and further south. The high probabilities associated with AR are located in the North Sea and the Mediterranean.

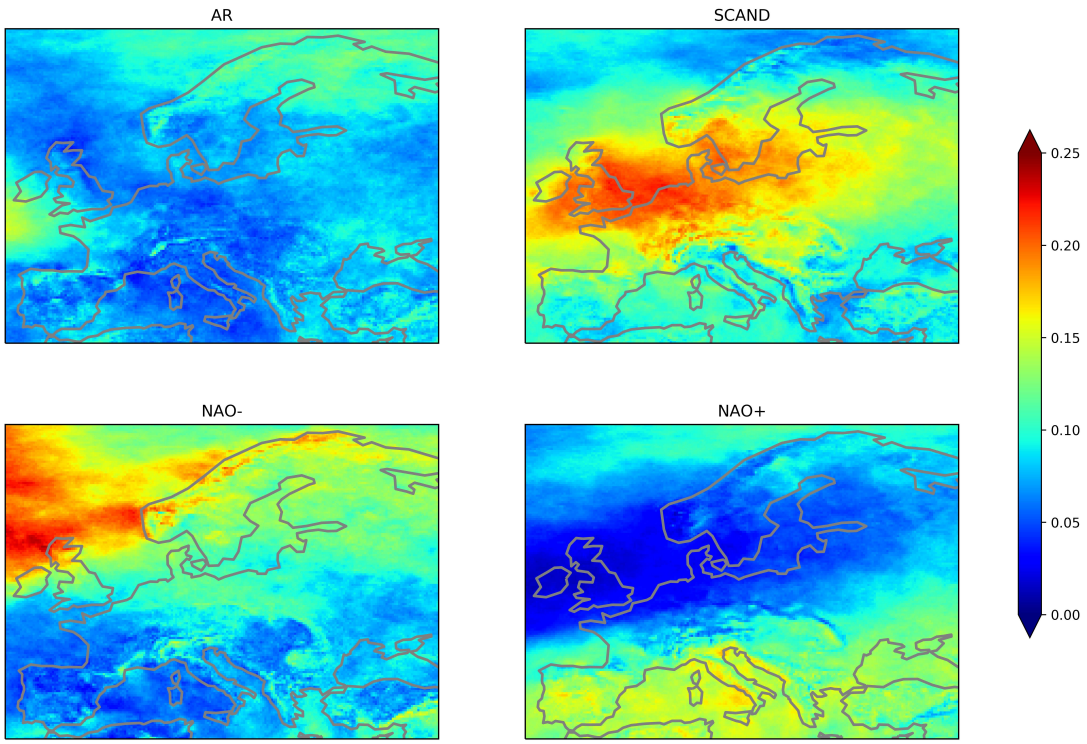

Figure 4.9: Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily maximum wind speed below the 10th percentile for each WR. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Wind speed In agreement with previous studies, Figure 4.9 indicates that blocking conditions lead to high probabilities of low wind speeds. Again, the spatial distribution of high probabilities associated with each regime is different. The SCAND pattern presents a maximum ($p_{X_{(SCAND)}} = 0.2265$) centered in the North Sea with high values extending across Northern Europe. This pattern presents, on average, the highest probabilities among all the WRs, $p_{\mu_{(SCAND)}} = 0.1246$, but also the lowest value of the four WRs ($p_{N_{(SCAND)}} = 0.0493$). The maximum probability over all regimes is registered during NAO- ($p_{X_{(NAO-)}} = 0.2400$). During NAO-, the maximum is located in the North Atlantic, with high values extending to Scandinavia, and low values in the south. An opposite gradient is inherent to the NAO+, with the lowest values in Northern Europe and Scandinavia. On average, probabilities of low wind speeds during NAO+ are $p_{\mu_{(NAO+)}} = 0.0862$. Meanwhile, the AR pattern does not show any particular signal and has the lowest averaged probabilities, $p_{\mu_{(AR)}} = 0.0797$.

Comparing Figures 4.8 and 4.9 shows high of low temperature and low wind speed events are located in approximately the same geographical areas, as an opposition to the AR and the

NAO+. Moreover, departures from the 10^{th} percentile are smaller for low wind speeds than for low temperatures (notice the different color bars).

Seasonality of marginal probabilities by regime

It has been shown above that the observed frequencies of low temperatures and low wind speeds have a pronounced seasonal behaviour and that the WRs proposed by Cassou are associated with specific patterns during the winter season. In this section, both results are combined by computing the seasonal empirical probabilities relative to each WR.

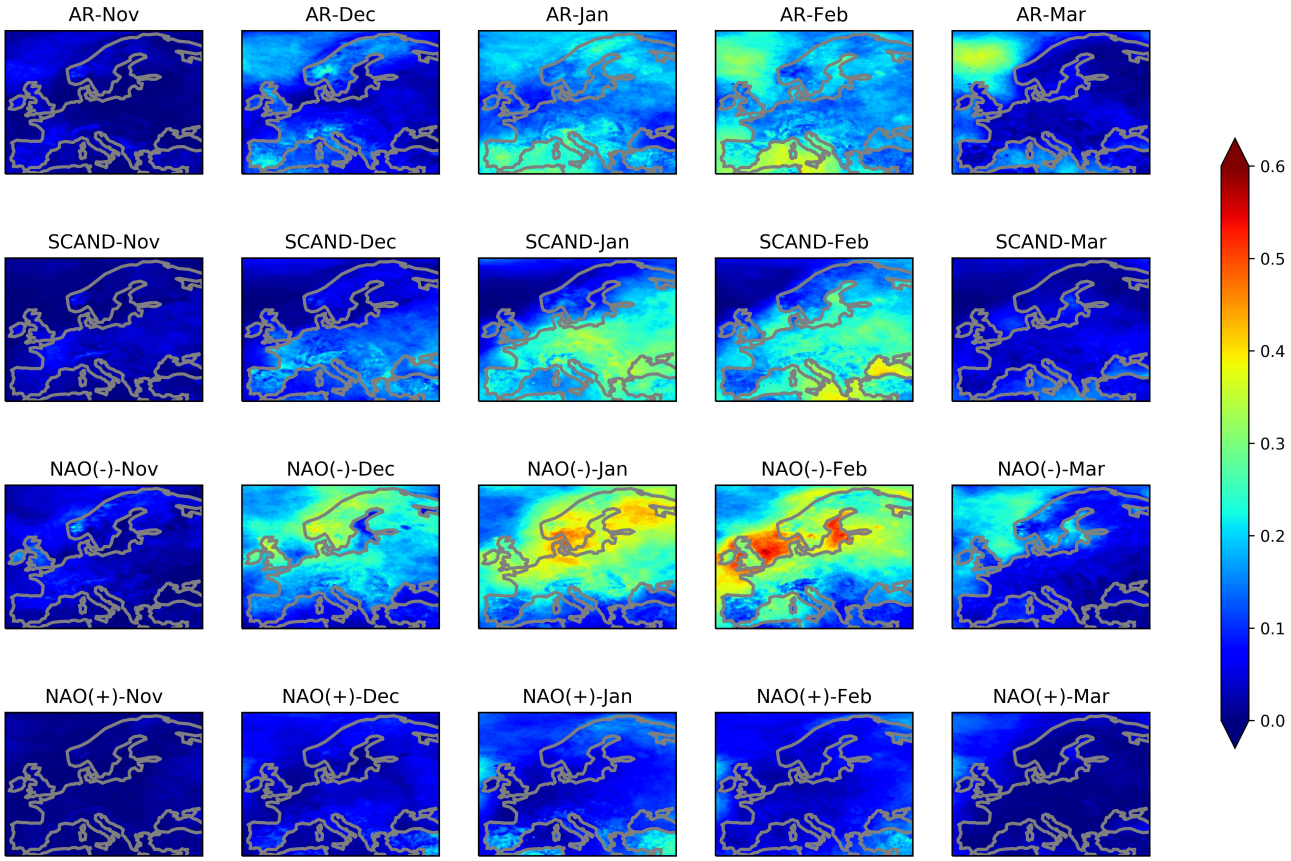

Figure 4.10: Marginal probabilities of low temperature events. Colours show the probabilities of occurrence of low temperature events computed as the proportion of days with the daily minimum temperature below the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Temperature On average, the probabilities of occurrence of daily minimum temperatures below the 10^{th} percentile increase during the coldest months, reaching a peak in February, as shown in Figure 4.10 and Table 4.5, except for the NAO+ that reaches the highest values in January. The maximum probabilities over all locations occur during the same months (see Table

Table 4.4: Maximum empirical marginal probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.1316	0.1262	0.2294	0.0756
Dec	0.2681	0.2628	0.4180	0.2500
Jan	0.3644	0.3836	0.5280	0.2949
Feb	0.4074	0.4453	0.5789	0.2515
Mar	0.3745	0.1969	0.3004	0.1976

Table 4.5: Mean empirical marginal probabilities of low temperature events p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.0196	0.0198	0.0375	0.0084
Dec	0.0915	0.0887	0.1834	0.0492
Jan	0.1721	0.1723	0.2538	0.1834
Feb	0.1997	0.1859	0.2618	0.0699
Mar	0.0196	0.0501	0.1001	0.0290

4.4). The NAO- is associated with the highest probabilities over major parts of Europe. It has a maximum located in the region formed by the North Sea and Baltic Sea, $p_{X_{(NAO-)}} = 0.5789$. In contrast, the lowest values, i.e., low probability of low daily minimum temperatures, are associated with warmer than normal temperatures during NAO+. Probabilities are also high in January and February during the SCAND over Central and Southwestern Europe, and during the AR in the Mediterranean and the Norwegian Sea. These two patterns have, on average, similar probability values, although located in different regions. The center of high probabilities located

over the Norwegian Sea, associated with the AR, intensifies in March.

In summary, the spatial distribution of probabilities of low temperature events has a strong seasonality at a regime level. NAO- has the strongest signal, whereas the opposite yields for NAO+. High probabilities can be expected from December to February, and even in March over the ocean during NAO-, SCAND, and AR.

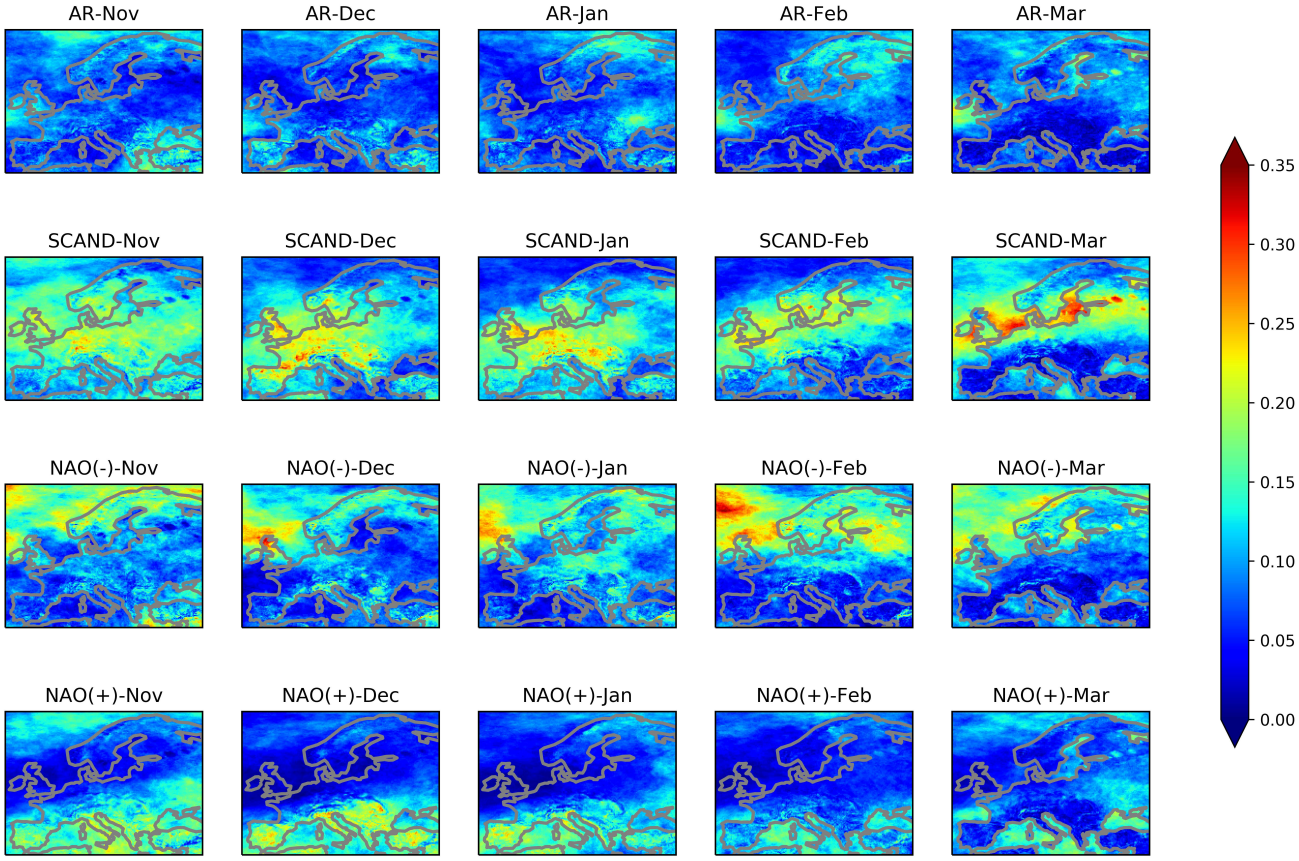

Figure 4.11: Marginal probabilities of low wind speed events. Colours show the probabilities of occurrence of low wind speed events computed as the proportion of days with the daily maximum wind speed the 10^{th} *percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

Wind speed The probability of occurrence of low wind speed events shows smaller seasonal variations than low temperatures for all the patterns but the SCAND (see Figure 4.11 and Tables 4.6 and 4.7). Although the highest maximum value is observed in February under NAO- conditions $p_{X_{(NAO-,Feb)}} = 0.3526$, on average, the probabilities of low wind speed events are higher under SCAND conditions for all the months of the season. As numerous studies describe, these two blocking patterns are associated with weak 10m winds. Hence, it is no surprise that

Table 4.6: Maximum empirical marginal probabilities of low wind speed events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.2218	0.3060	0.3059	0.2773
Dec	0.2255	0.3494	0.3484	0.3090
Jan	0.2089	0.2956	0.3494	0.2769
Feb	0.1975	0.2920	0.3526	0.2101
Mar	0.2383	0.3406	0.2767	0.2212

Table 4.7: Mean empirical marginal probabilities of low wind speed events (p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.0856	0.1372	0.1160	0.1037
Dec	0.0778	0.1249	0.1014	0.0849
Jan	0.0800	0.1217	0.1113	0.0848
Feb	0.0776	0.1131	0.1198	0.0743
Mar	0.0768	0.1248	0.1088	0.0826

the probabilities of low wind speed events are high during these patterns. The spatial distribution is, however, different. High probabilities are located further north during NAO- with respect to SCAND, as happens with the probabilities for low temperature events (see Figure 4.10. The strongest north-south gradients are observed in March, during SCAND; probabilities are high across Central and Northern Europe this month, with a maximum located over the sea ($p_{X_{(NAO-,Mar)}} = 0.3406$) and low in southern Europe ($p_{N_{(NAO-,Mar)}} = 0.0000$). It is also evident that probabilities are higher over water bodies than over land in March. Contrarily,

the continent is affected by anticyclonic conditions from November to January, being February a transition month. The maximum probability observed during this cluster occurs in December, $p_{X_{(SCAND,Dec)}} = 0.3494$. On average, November is the month with the highest probabilities, $p_{\mu_{(SCAND,Nov)}} = 0.1372$. Although this month does not show the highest values, they are relatively high over large parts of the domain.

Low wind speed events are also expected during NAO- at latitudes higher than 60°N, in particular, during February in the Atlantic and North Sea ($P_{X_{(NAO-,Feb)}} = 0.3526$). This pattern shows very small seasonal variations, especially the spatial structure. Average values range from $p_{\mu\,(NAO-,Dec)} = 0.1014$ in December to $p_{\mu_{(NAO-,Feb)}} = 0.1198$ in February. On the other hand, the spatial distribution of the NAO+ is the opposite of the negative phase and weaker. The highest probabilities during NAO+ conditions occur in the south from November to January. The highest frequency of low wind speed events peaks in December ($p_{X_{(NAO+,Dec)}} = 0.3090$) but, on average, more events are observed in November ($p_{\mu(NAO+,NOV)} = 0.1037$). The lowest frequency of events is observed during the AR. Wind speed probabilities are, in general, lower than temperature probabilities.

4.3.2 Joint probabilities

The main goal of this thesis is to model the co-occurrence of low temperature and low wind speed events in the light of the WRs for each month of the winter season. Historical joint distributions are described in this section.

Seasonality of joint probabilities

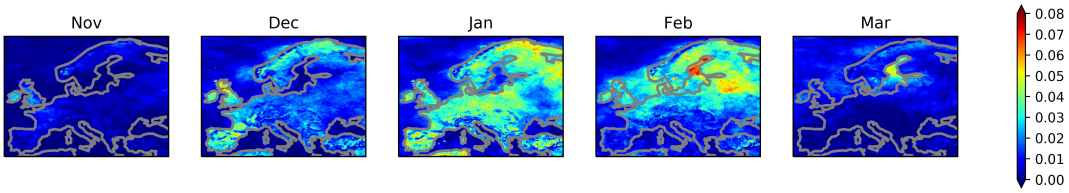

Figure 4.12: Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10^{th} percentile for each month. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

The seasonality of the combined events is summarized in figure 4.12. Joint events of low temperatures and low wind speeds are frequently observed across the continent during the cold-

est winter months. Already in December, some high probabilities are observed in Western Europe and Scandinavia ($p_{X_{(Dec)}} = 0.0793$, $p_{\mu_{(Dec)}} = 0.0132$), where low temperatures coincide with low wind speeds. In January, the entire continent is affected by an enhanced frequency of co-occurrence of extremely low temperatures and low wind speeds ($p_{X_{(Jan)}} = 0.0741$, $p_{\mu_{(Jan)}} = 0.0218$), but, in February, the highest probabilities are located over the Baltic Sea and Northwestern Europe ($p_{X_{(Feb)}} = 0.0775$, $p_{\mu_{(Feb)}} = 0.0202$). This month, Northern Europe and Scandinavia (both landmasses and water bodies) are hit by relatively frequent joint events. In March, only the Baltic sea is governed by frequent compound events, whereas the rest of the domain experiences very low probabilities. Very low probabilities are observed in the month of November ($p_{\mu_{(Nov)}} = 0.0030$).

Joint probabilities computed by regime

Empirical joint probabilities were also analyzed on a WR level for the entire season (NDJFM), showing the importance of considering the WRs in the study of compound events.

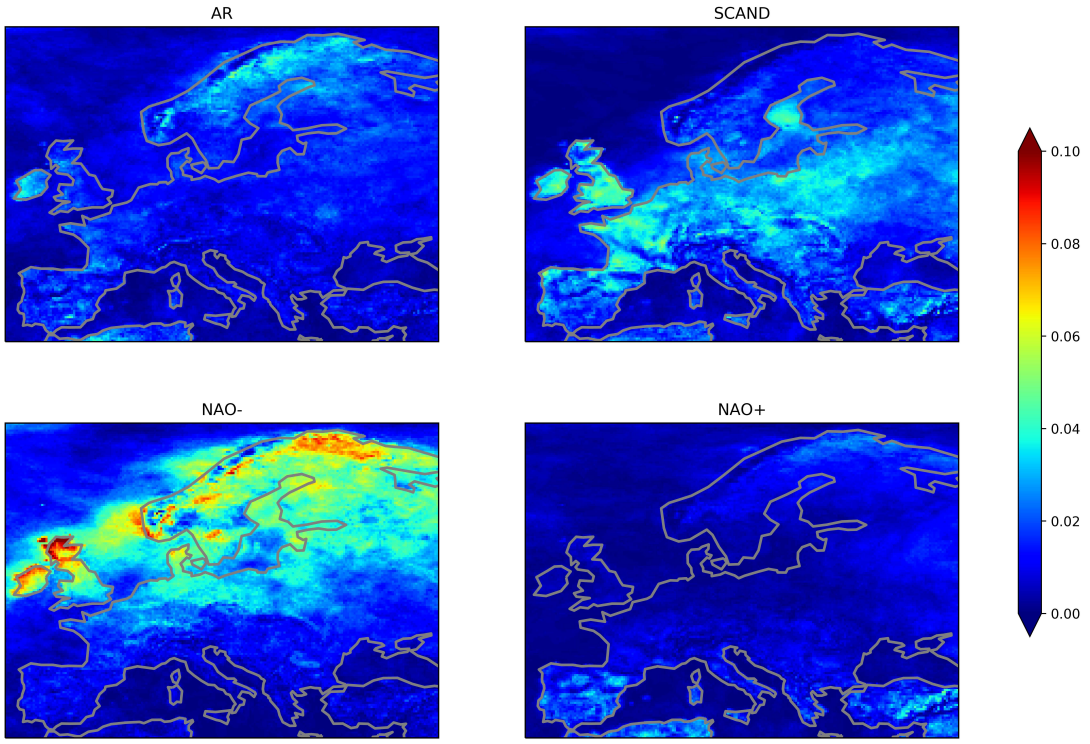

Figure 4.13: Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table 4.8: Maximum empirical joint probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.0489	0.0694	0.1000	0.0224
Dec	0.1106	0.1218	0.1885	0.0899
Jan	0.1156	0.1478	0.1729	0.1205
Feb	0.1070	0.1387	0.1947	0.0858
Mar	0.0426	0.0906	0.1107	0.0472

Figure 4.13 reveals that the days on which the joint effect is most prominent are classified as NAO-. For this pattern, the highest observed joint probabilities during the winter season are $p_{X_{(NAO-)}} = 0.1120$, and the average is $p_{\mu(NAO-)} = 0.0231$, both double as high as for the other regimes. Northern Europe and Scandinavia are the regions affected by compound events, over land, but also over the sea. The SCAND pattern also shows relatively high probabilities of joint events over an extended area in Central and Northern Europe, but not as prominent as the NAO-. The maximum and mean values observed are $p_{X_{(SCAND)}} = 0.0584$ and $p_{\mu_{(SCAND)}} = 0.0150$ respectively. During this regime, very low probabilities are observed in the Atlantic and the Mediterranean Sea. The AR pattern consists of high probabilities in Scandinavia and Ireland, whereas the NAO+ is not affected by the co-occurrence of low temperature and low wind speed events.

As the marginal probabilities figures indicate, high probabilities of low temperature and low wind speed events are located over the same area; thus, joint probabilities are high for these blocking patterns. It is worth mentioning that the meteorological conditions impose a more substantial effect on the joint probabilities than the seasonality, and that the color bars are truncated at different levels.

Seasonality of joint probabilities by regime

Finally, the combined seasonal and effect on the historical joint distributions is reported.

Figure 4.14 displays the seasonality of the empirical joint probabilities for each WR. As

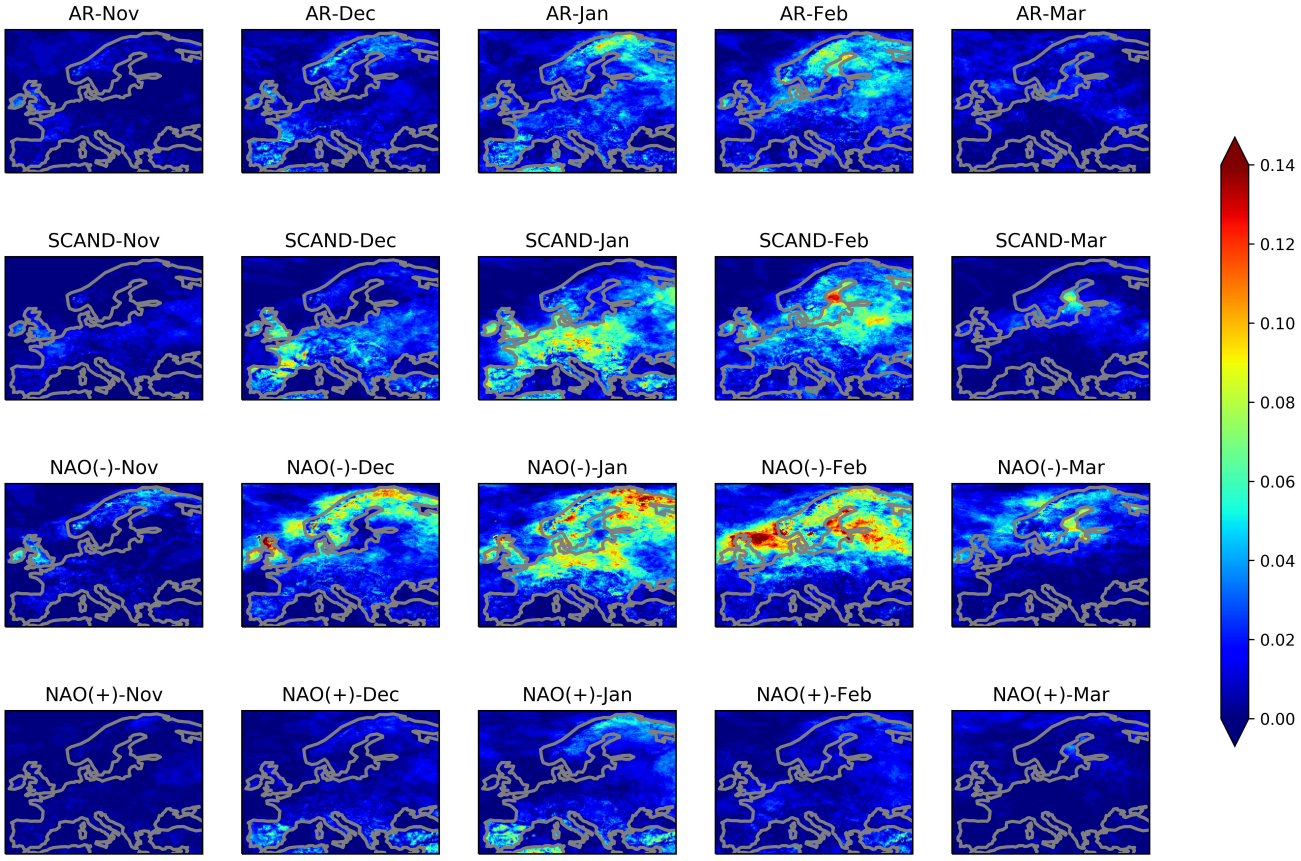

Figure 4.14: Joint probabilities of low temperature and low wind speed events. Colours show the probabilities of occurrence of low temperature and low wind speed events computed as the proportion of days with daily minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

expected, joint probabilities are high when and where both marginals are high. The strongest signal occurs during NAO-, from December to March, with a maximum in the coldest months of the year, in Northern Europe and Scandinavia. The highest relative frequency of events is observed in February associated with the NAO- pattern, $p_{X_{(Feb,NAO-)}} = 0.1947$ and $p_{\mu_{(Feb,NAO-)}} = 0.0357$. The other three patterns reach the highest probabilities in January (see Tables 4.8 and 4.9. The SCAND has the second most prominent pattern in the winter season, with a region of high probabilities that covers most of the continent, extending from southwest to northeast, in January, when it reaches the maximum ($p_{X_{(Jan,SCAND)}} = 0.1478$, $p_{\mu_{(Jan,SCAND)}} = 0.0270$). In December, the co-occurrence of low temperature and low wind speed events is favoured in northern Spain, western France, and Britain; these high probabilities move northwest and are centered over the Baltic Sea in February. High probabilities characterize the AR in Northwestern Eu-

Table 4.9: Mean empirical joint probabilities of low temperature events (p_μ). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.0027	0.0037	0.0055	0.0013
Dec	0.0093	0.0145	0.0246	0.0067
Jan	0.0178	0.0270	0.0348	0.0126
Feb	0.0203	0.0246	0.0357	0.0079
Mar	0.0068	0.0065	0.0143	0.0028

rope and Scandinavia in January and February. As expected, NAO+, the pattern associated with warm anomalies across Europe, has the lowest probabilities for joint events.

The four WRs have in common that November and March are the months with the lowest probabilities and that a strong seasonality is reflected both in the number of occurrences and in the geographical area affected by compound events. As blocking situations lead to high marginal low temperatures and low wind speeds in approximately the same regions, high joint probabilities are also expected during these regimes. Despite the fact that blocking patterns are associated with the strongest signals, the other regimes are also affected by high probabilities in some locations during the coldest months. As seen in Figure 4.13, WRs play an essential role in predicting joint probabilities, but Figure 4.14 shows that seasonal variations are big and should also be considered when modeling the multivariate distributions. Furthermore, joint probabilities calculated by aggregating the data by month and regime (Figure 4.14) are higher than when aggregated only by month (Figure 4.12) or only by regime (Figure 4.13).

4.3.3 Independence

To conclude the study of the empirical distributions, the dependence of temperature and wind speed events was assessed, and correlation values were computed. Events are dependent if they affect each other.

By comparing the probabilities from the joint distributions with the product of the marginal probabilities, the variables are proven to be dependent. The events are independent if and only if

these two quantities are equal (Equation 2.2. The following figures show the difference between the joint probabilities and the product of the marginals. This magnitude quantifies the strength of the association between the events. Positive values indicate that low temperature events favour low wind speed events, or, equivalently, that low wind speed favour low temperature events. The opposite holds for negative values. This is, the events are negatively correlated, and the presence of one event reduces the odds of the other event.

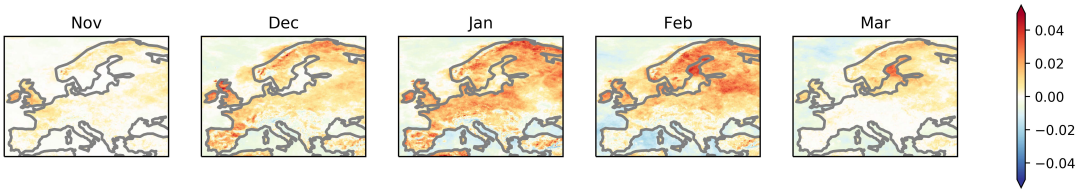

Figure 4.15: Seasonality of difference between the joint probabilities and the product of the marginal probabilities. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).).

The illustration of monthly differences between joint and product of marginal probabilities suggests that the dependence is stronger during the coldest months when marginal probabilities of low temperature events are highest (Figure 4.15). The biggest positive difference is observed in January, $\Delta_+ = 0.0481$. The biggest negative difference is also registered in January, $\Delta_- = -0.0274$. In general, the continent exhibits positive differences in areas where the empirical joint probabilities are high. This suggests that more events occur where the dependence is strongest. In contrast, the maritime regions are characterized by negative or close to zero values.

As Figure 4.15, Figure 4.16 shows that the differences between the joint probabilities and the product of marginals are positive across the continent when data is aggregated by WRs. The biggest differences occur during NAO- over Northern Europe and Scandinavia ($\Delta_+ = 0.0645$), where both events co-occur with more frequency. The most negative values are also observed during NAO-, $\Delta_- = -0.0273$, but over the ocean.

The combined effect of WRs and seasonality can be appreciated in Figure 4.17. More often than not, the differences are larger where the joint probabilities are highest, i.e., during the coldest months and under blocking conditions. In particular, the NAO- is associated with large differences, both positive and negative. This regime presents positive differences bigger than 0.9 from December to February; the biggest negative difference is observed in February and is $\Delta_- = -0.0515$. The differences are close to zero during the transition months, November and

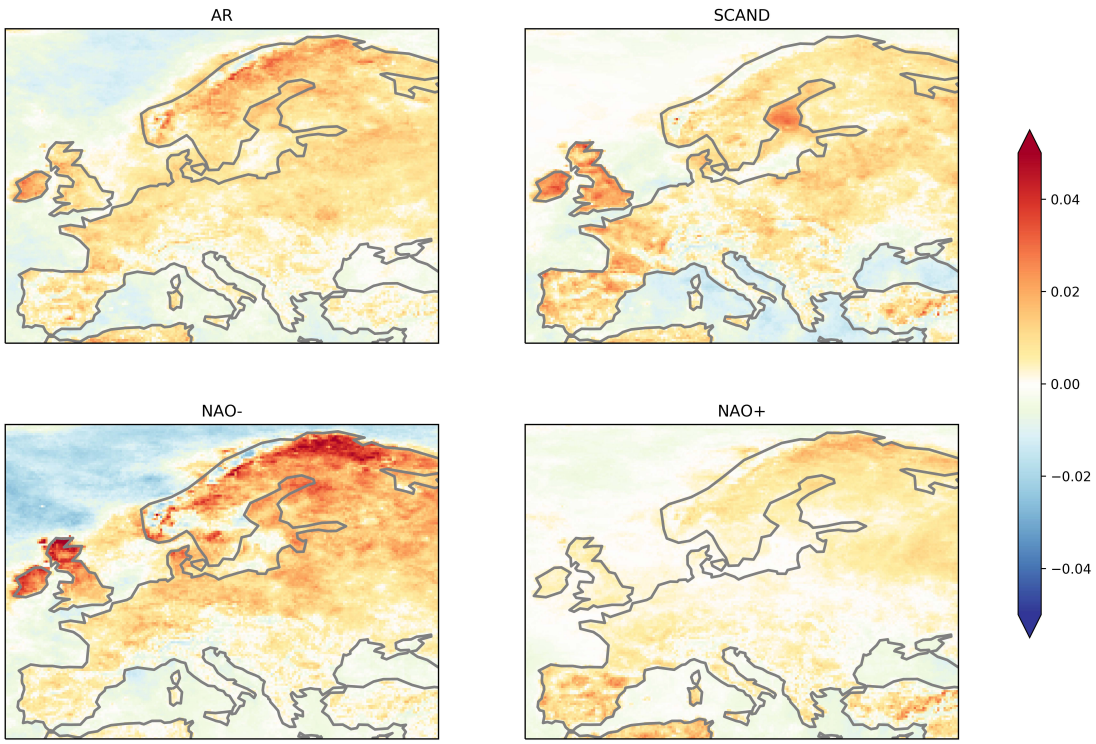

Figure 4.16: Difference between joint probabilities and product of marginals for each WR. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

March, for all the regimes.

Complementing these results, the chi-square test was performed at a 10% significance level, confirming that the events are dependent on most of the European domain. The results are available in the Appendix (8).

4.4 Bivariate analysis with Gaussian copula theory

4.4.1 Estimation of the copula parameters

Dependence and correlations are often treated as synonyms, but dependence is more general than correlation. If the variables are independent, the correlation is 0, but the converse is not true. The average posterior samples of the correlation matrix for low temperature and low wind speed events for each WR are computed as in D. Hoff, 2007 and shown in Figure 4.18. These correlation values are the parameters of the copula function employed to model the bivariate distributions.

The four WRs have in common that the correlations between low temperature and low wind

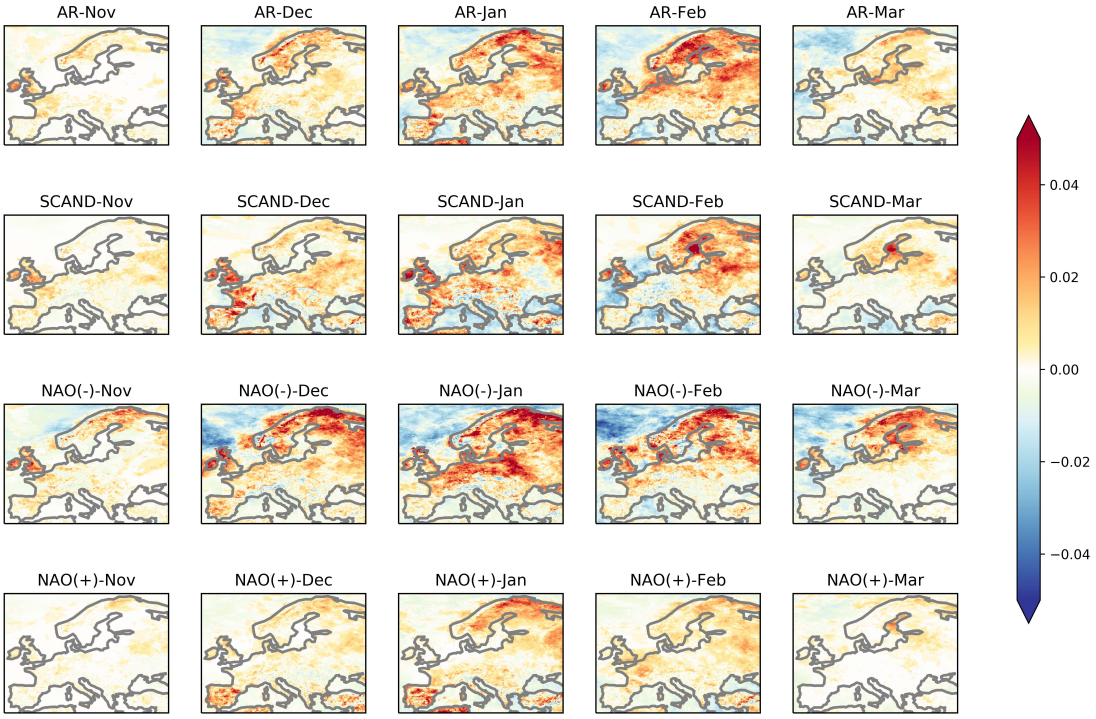

Figure 4.17: Seasonal difference between joint probabilities and product of marginals for each WR. Red (blue) tones indicate that the joint probabilities are greater (less) than the product of marginals. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table 4.10: Maximum (ρ_X), minimum (ρ_N), and average (ρ_μ) correlations for each regime. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

WR	ρ_X	ρ_N	ρ_μ
AR	0.71	-0.70	0.23
SCAND	0.67	-0.69	0.18
NAO-	0.63	-0.67	0.21
NAO+	0.70	-0.64	0.19

speed events are positive on mainland Europe, except some high locations, such as the Alps and the mountains in Norway. In contrast, over the sea (except the North Sea and the Baltic Sea), low temperatures and low wind speeds are, in general, anticorrelated.

The highest correlations are observed during the AR (the maximum is $_{XAR} = 0.71$). The

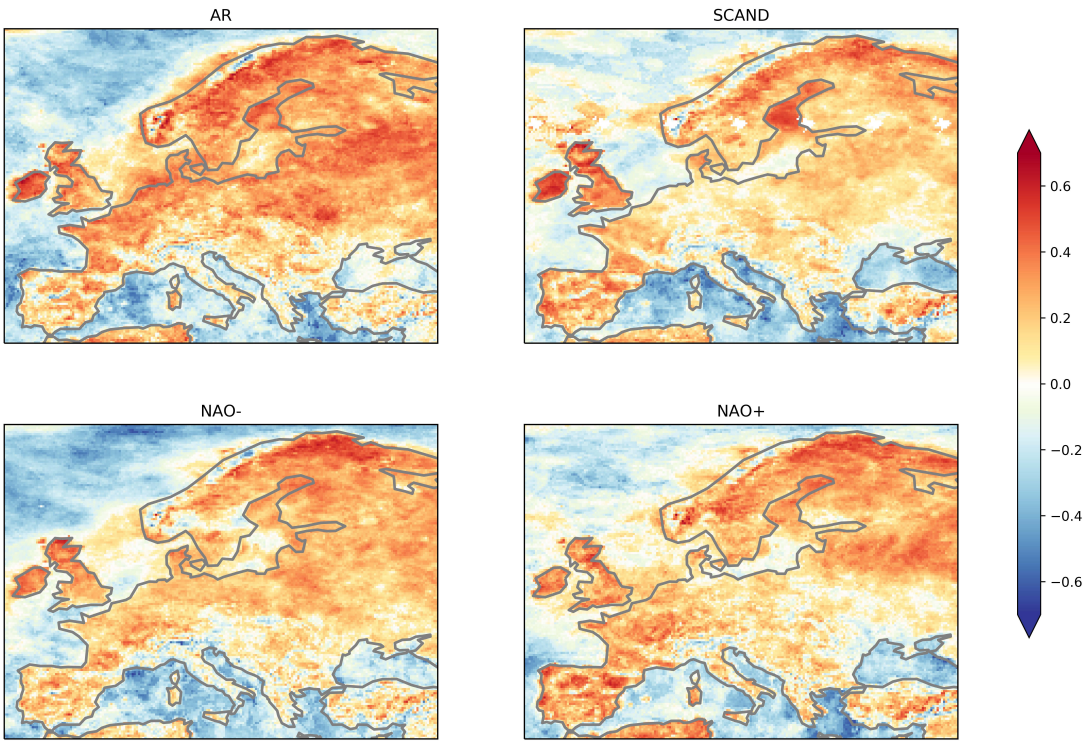

Figure 4.18: Correlations (ρ) between low temperature events and low wind speed events for the four WRs. The threshold is defined as the below the 10^{th} percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

minimum and the average are also the highest compared to the other WRs (see Table 4.10) It presents a meridional gradient over the continent. Also, the North Sea and the Baltic Sea present the highest correlations during this pattern. The correlations are lower over Central and Eastern Europe during SCAND and NAO+, although they are higher in the Iberian Peninsula. Over the sea, in particular, over deep waters, the variables are anticorrelated.

Add table with average of positive values?

The seasonal changes of the correlations can be observed in Figure 4.19. November is the warmest month of the winter season, with less low temperature events than other months. Consequently, the correlations cannot be computed at every location. The most prominent contrasts between positive and negative correlations are manifested during February and March under AR conditions. Also, during the AR, the positive correlations over Scandinavia are exceptionally high in January. Events occurring under SCAND conditions are positively correlated in Eastern Europe, with a maximum by the end of the season, and in Britain. The NAO- presents a stationary maximum over Scandinavia. Contrarily, the maximum located over Northern Europe is strongest in January. NAO+ shows a strong seasonality; in January, the maximum is

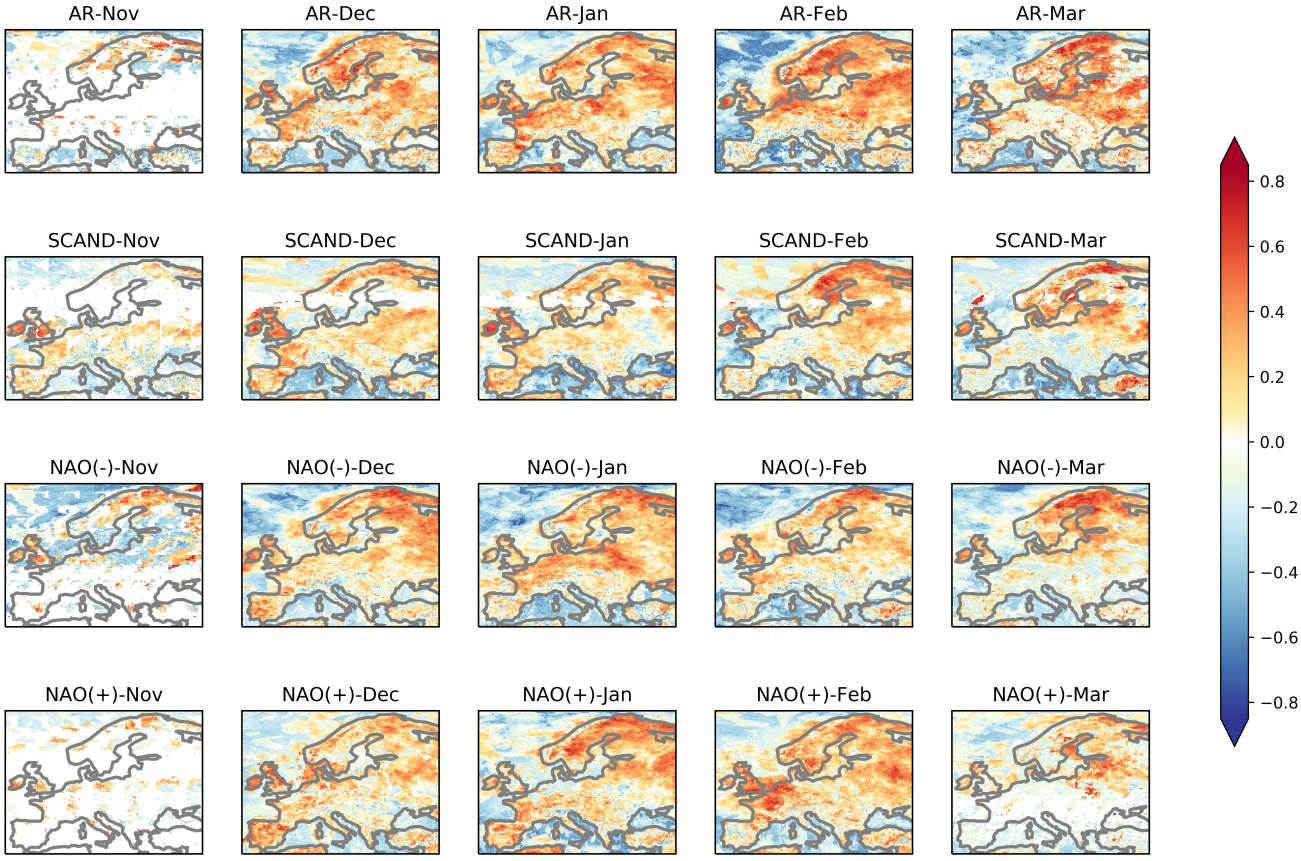

Figure 4.19: Seasonality of correlations (ρ) between low temperature events and low wind speed events for the four WRs. The threshold is defined as the below the 10^{th} percentile. Locations with no low temperature event during the selected period have no colour. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

located over Scandinavia, whereas in February, the maximum is located over Northern Europe and Britain. Correlations over the Iberian Peninsula are highest in January and February. Notice that the correlations are higher when computed for each month separatetly than for the entire season (Figure 4.18).

4.4.2 Models for marginal distributions

A logistic regression model was proposed to determine the marginal probabilities of occurrence of low minimum daily surface temperature events and low maximum daily wind speed events on each grid box. The univariate distributions, together with the correlations, are the inputs to the copula model.

Three different models are compared, a climatology model that only considers the month in which the events occur; a regime model that only employs the four WRs as predictors; and an interaction model that takes into account both sets of predictors. Given that only two sets

of binary predictors are involved in the models, the modeled marginal probabilities show no difference with respect to the empirical probabilities shown in the previous section. However, the goal of this thesis is to provide a methodology for computing the multivariate probabilities, in which case the marginal probabilities can be modeled with more complex models that the provided here.

The results obtained with the logistic regression models are provided in the appendix, but, as already mentioned, the figures resemble those of the odds ratios.

It has been shown in the figures for the empirical distributions that the WRs play an important role in modeling the univariate distributions. The BSS computed for the modeled marginal distribution with the interaction model described in section 3 compared to the climatology model (that does not use any information about the WRs) is a measure of the relative performance. The predicted marginal distribution with the logistic regression approach was obtained with the cross-validation method, more specifically, by leaving one year out. In other words, for each year, the probabilities are computed by fitting the model to the rest of the sample.

Regime model

The BSS were also computed for each WR, and the significance was assessed with a permutation test at the 10% level. For this purpose, the regime model consisting of a set of binary predictors for the regimes (described in the Methodology section 3) was employed. Positive values indicate that the model regime model, which only considers information regarding the WRs, scores better than the climatology model.

Temperature Significant scores for the low temperature marginals for each regime are represented in figure 4.20. The SCAND and the NAO+ score highest in regions where probabilities of low temperature events are low. The NAO- scores high over Britain and Scandinavia, and over the Mediterranean. The AR shows a small improvement compared to the climatology in the North Sea.

Wind speed The NAO+ scores also highest when predicting low wind speed events, where the probabilities are lowest. The other regimes have comparable scores, in regions where probabilities are low and where probabilities are high; for instance, the SCAND scores high over Northern Europe and the NAO- scores high in the North Sea.

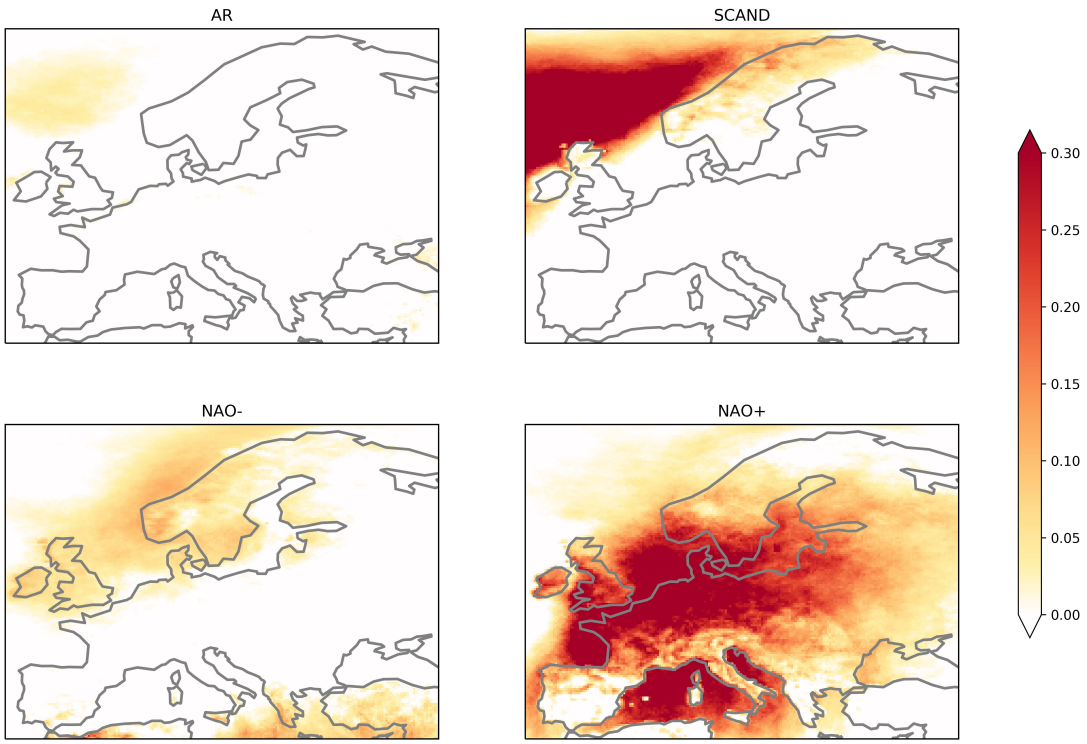

Figure 4.20: Significant BSS for minimum daily surface temperature for each WRs. A monthly seasonal model was used as a reference. Colours show significant BSS values at the 10% level that are not zero, and the color bar is truncated at BSS = 0 and BSS = 0.7. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Interaction model

The interaction model consists of two sets of predictors, one for the months of the winter season and one for the WRs. Here, the probabilities computed with the interaction model are compared to the climatology model, in terms of the BSS.

Temperature The BSS were also computed for each WR and month combination, and the significance was assessed with a permutation test at the 10% level. Figure 4.22 shows a significant improvement with respect to the climatology (values are positive) across large geographical areas. Over land, NAO+ is the regime with the highest BSS, with an average BSS ranging from 0.16 to 0.26. The SCAND scores highest over the North Atlantic, reaching values higher than 0.9 in every month. These regions are characterized by very low probabilities for cold events (Figure C.11. Days classified as NAO- score highest in the region conformed by Scandinavia, Britain, and the North Sea, during the coldest months (see Table 4.11). The AR shows a great intraseasonal variation, with high significant values over the sea and significant values over land

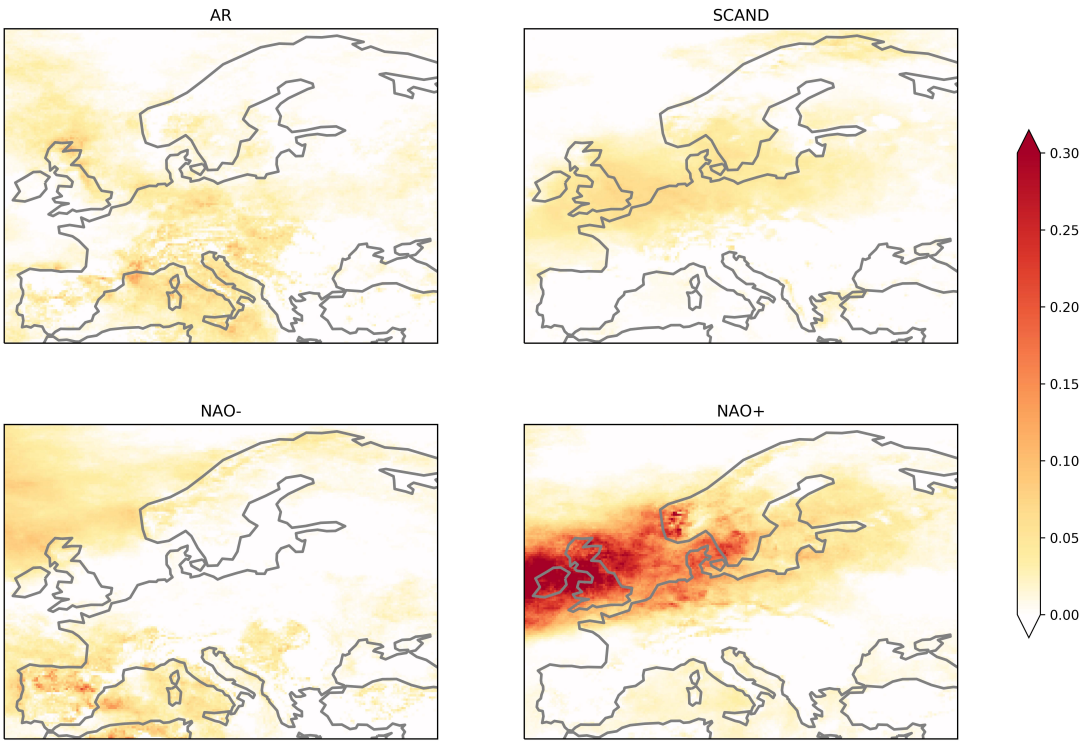

Figure 4.21: As figure 4.20 but for maximum daily 10m wind speeds. Colours show significant BSS values at the 10% level, and the color bar is truncated at BSS = 0 and BSS = 0.3. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

in February and March. In general, it is observed that the regions affected by high probabilities of occurrence of low temperature events (see Figure 4.10 have a positive and significant BSS.

Wind speed As for the marginal temperature probabilities, the highest significant BSS values are associated with NAO+ conditions during the coldest months, but only in a restricted area in Northwestern Europe (see Figure 4.23 and Table 4.12. During NAO-, BSS are positive across the Mediterranean and in the North Atlantic, coinciding with the regions where probabilities also are high. The maximum score is $BSS_{X,Jan,NAO-} = 0.9745$. Meanwhile, during SCAND, the significant values are located over Northern and Central Europe from November to January, and over Northern Europe and Scandinavia by the end of the season, reaching the highest during the coldest months. In these regions, the probability of occurrence of weak winds is also relatively high. AR conditions lead to no significant pattern for the BSS. In general, all the significant values are positive, meaning that the interaction model performs better than the climatology model, i.e., the WRs are important predictors for the marginal probabilities. Furthermore, the areas affected by high probability of occurrence of weak wind events (see Figure 4.11 present

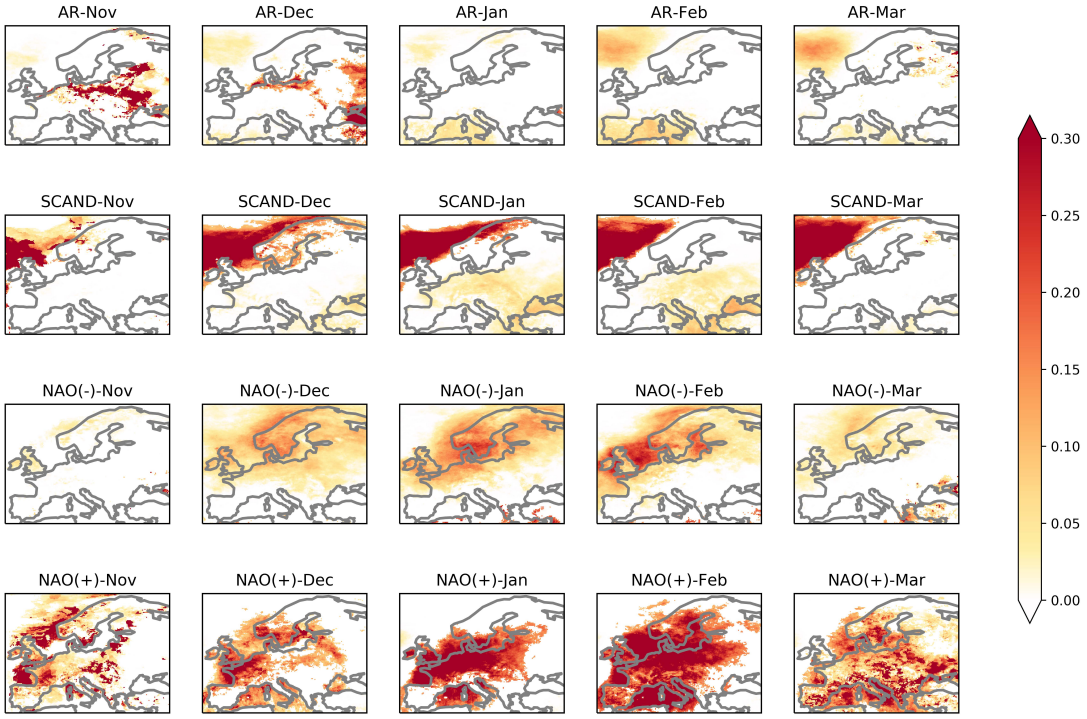

Figure 4.22: Seasonality of ignificant BSS for minimum daily surface temperature for each WRs. A monthly seasonal model was used as a reference. Colours show significant BSS values at the 10% level that are not zero, and the color bar is truncated at BSS = 0 and BSS = 0.3. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table 4.11: Maximum significant BSS for low temperature events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.8469	0.9027	0.5574	0.9187
Dec	0.9661	0.9853	0.3022	0.9441
Jan	0.3358	0.9887	0.4666	0.9948
Feb	0.1397	0.9885	0.3304	0.8034
Mar	0.6740	0.9893	0.8429	0.9635

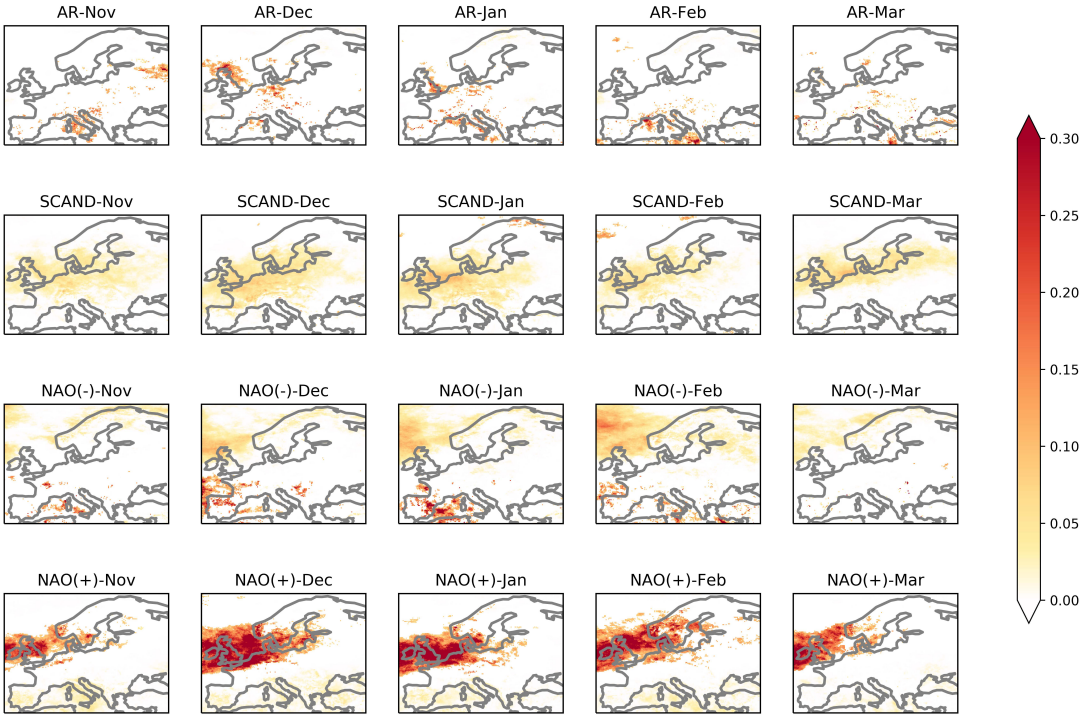

Figure 4.23: As figure 4.22 but for maximum daily 10m wind speeds. Colours show significant BSS values at the 10% level, and the color bar is truncated at BSS = 0 and BSS = 0.3. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

also significant BSS.

The main conclusion from this section is that the WRs provide useful information for the prediction of the marginal distributions, but the seasonality is also an important factor. For both variables, it can be observed that the interaction model, which considers both the months and the regimes, scores better than the regime model, which only considers the regimes.

4.5 Joint probabilities

The copula function makes it possible to generate a joint distribution from arbitrary marginal probabilities. The joint probabilities for daily wind speeds and temperatures exceeding the threshold defined by the 10^{th} percentile of each variable are exhibited in this section.

Figure 4.24 is the result of applying the copula function to the uniform distribution derived from the marginal probabilities modeled with logistic regressions. This methodology reproduces the results obtained for the empirical distributions because the marginal distributions were modeled using only sets of binary predictors, thus obtaining the observed marginal probabilities. As mentioned before in this document, the aim is to establish a flexible methodology for using Gaussian copulas to model the joint probabilities. Figure 4.24 validates the use of the

Table 4.12: Maximum significant BSS for low wind speed events. Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.5407	0.1142	0.3772	0.4761
Dec	0.5120	0.1486	0.4877	0.9849
Jan	0.3966	0.2521	0.9745	0.7880
Feb	0.5991	0.2067	0.9008	0.7919
Mar	0.8024	0.1337	0.6980	0.6526

methodology proposed.

Finally, the importance of considering the associations between the variables when modeling the joint distributions is documented. For this purpose, the BSS were computed for a Gaussian model with dependence (modeled as in D. Hoff, 2007) relative to a model that assumes that the events are independent.

Significant BSS computed for each WR are exhibited in figure 4.25. The regimes with the highest scores are NAO- and AR. Both regimes have positive and significant scores in Scandinavia and Northern Europe, indicating that the model accounting for associations between low temperature and low wind speed events performs better than the independent model. The SCAND presents a noisier signal in Central and Northern Europe, although large continental areas have a positive skill as well. The dependent model also scores better during NAO+ but values are smaller than for the other regimes. The BSS are negative in regions were the dependent model doe not improve the independent model. For instance, in the North Sea, during SCAND conditions, or in Central Europe during NAO-. These are also regions where joint probabilities are low (see figure 4.13), i.e., where chances of both events occurring simultaneously are low.

The seasonality of significant BSS is illustrated in Figure 4.26. The transition months and the NAO+ regimes are excluded because they register very few low temperature events. An important intermonthly variation is observed, which explains why the scores computed on a monthly basis are higher than those for the entire season (Figure 4.25). The seasonality of the

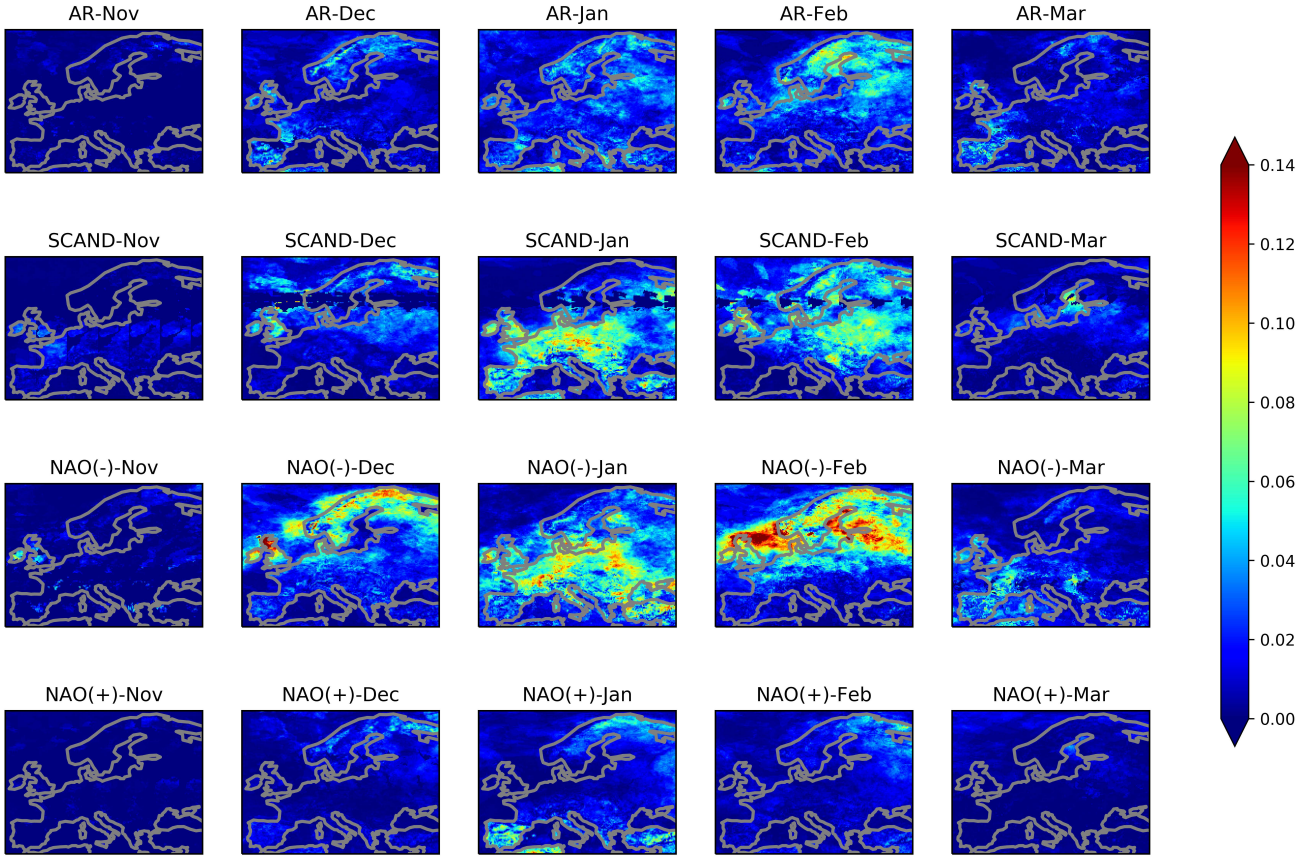

Figure 4.24: Seasonality of joint probabilities from the copula model for each WR. Colours show the probabilities of occurrence of low temperature and low wind speed events computed with the copula model for minimum temperatures and daily maximum wind speeds below the 10^{th} *percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

scores is in line with that observed for the join probabilities (see Figure 4.24), showing the potential of the Gaussian copula framework in the estimation of joint meteorological events and the importance of considering associations between variables when modeling joint distributions. However, the values are low, meaning that there is room for improvement, defining, for instance, a more complex model for the marginal probabilities.

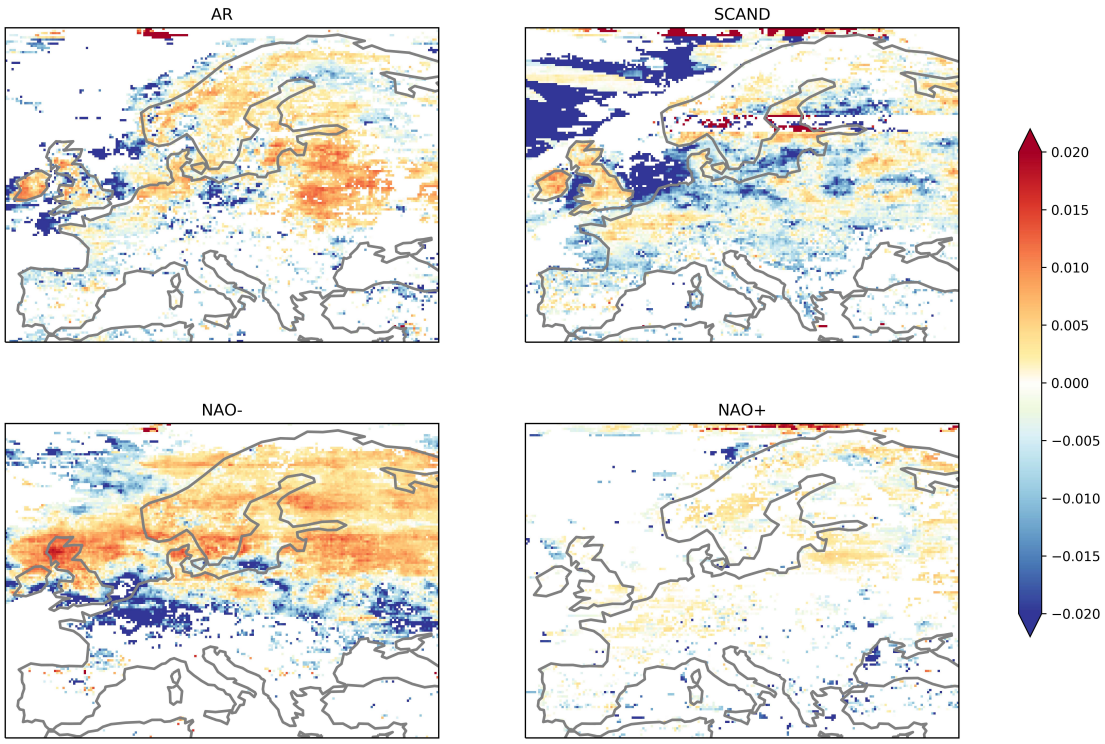

Figure 4.25: Significant BSS at the 10% level for joint probabilities modeled with Gaussian copulas and logistic regressions, aggregated by WRs. An independent model, with no correlation between the temperature and the wind speed events, was employed as the reference model. Colours show significant BSS values that are not zero, and the color bar is truncated at BSS $= 0.03$*. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

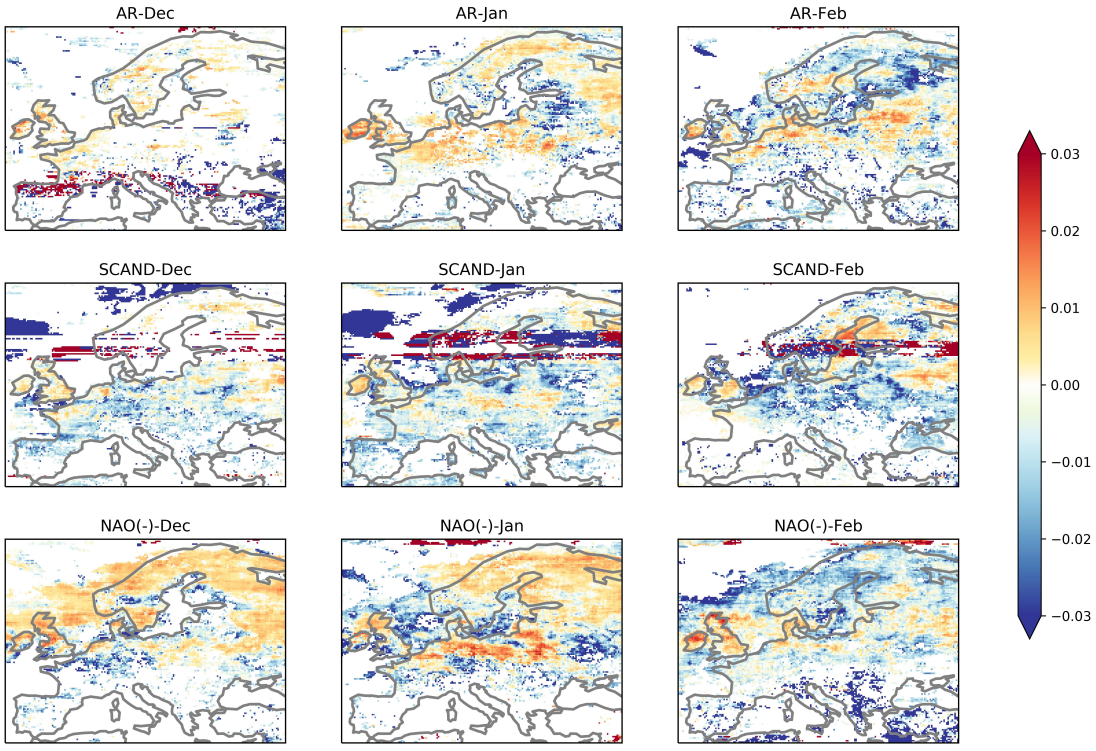

Figure 4.26: Significant BSS for joint probabilities modeled with Gaussian copulas and logistic regressions, aggregated by WRs (AR, SCAND, NAO-)and month (DJF). An independent model, with no correlation between the temperature and the wind speed events, was employed as the reference model. Colours show significant BSS values that are not zero, and the color bar is truncated at BSS $= 0.03$*. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

5 Discussion

This section summarises the findings and contributions made while comparing the results to those of older studies. Interesting research questions for future research are also discussed.

A framework for modeling monthly multivariate distributions of meteorological variables associated with low electricity production and high demand conditioned to the occurrence of large-scale atmospheric circulation is proposed in this thesis. Joint probabilities of low temperature and low wind speed events occurring during the extended boreal season (NDJFM) on the European domain are characterized. Atmospheric variability is represented by the Wintertime North Atlantic weather regimes. These regimes are computed using the k-means algorithm for $k = 4$ clusters, obtaining atmospheric circulation patterns that can be viewed as the positive and negative phases of the NAO, which are essentially a measure of the variability of the zonal flow over the North Atlantic; the positive phase of the SCAND, and the positive phase of the AR (R. Vautard, 1990, Hurrell et al., 2003; Cassou, 2008). Compared to other papers (e.g., Cassou, 2008; Wiel, Hannah C Bloomfield, et al., 2019), the mean circulation patterns derived for each WR do not show any considerable differences, although the period selected for the computation or the reanalysis data is different. For this work, the WRs were computed for the first 14 of Z500 from ERA5 for the period 1980-2017. The difference in the percentage of the number of days, representing the mean frequency of occurrence of each regime, is greatest for the NAO+ and the AR compared with the regimes computed by Cassou (2008) and Wiel, Hannah C Bloomfield, et al. (2019) respectively. In both cases, the difference is equal to 2%. The rest of the WRs show a maximum difference of 1%. Despite the difference in the percentage of days, all three studies agree that the most frequent pattern is the NAO+, followed by the SCAND.

For sake of comparison, mean average temperature and wind speed conditions were determined for each WR, replicating the results obtained by Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020) and Wiel, Hannah C Bloomfield, et al. (2019). This confirms the robustness of the methodology followed for the computation of the WRs, on which the rest of the analysis is based. Looking at the fields of mean meteorological conditions determined for each WR (Figures 4.1 and 4.2), it is evident that the impacts of the NAO- and the SCAND on near-surface variables are associated with cold and weak wind conditions, in line with previous studies. As a consequence, it is expected a higher frequency of events for

days classified in these two clusters. The study of observed odd ratios confirms this hypothesis. It expands our knowledge about meteorological events associated with the WRs by providing an insight into how many events we can expect given the defined threshold, instead of only focusing on the departures from average conditions.

It follows from the observed frequencies of joint events classified in each WR (Table 4.3) that the percentage of the total number of joint events is 1.29%, and the NAO- and the SCAND together count for 0.84%. These numbers represent average percentages over the entire grid, and should not be confused with local maximums. From the spatial distribution of joint events, manifested in figure 4.13, it is evident that high probabilities are constrained to restricted regions of the domain. Thus, much higher values are expected at specific locations. In the case of the NAO-, the geographical region that exhibits high observed frequencies of joint events is restricted to Northern Europe and Scandinavia. At some locations, the frequency of observed joint events exceeds 10% of the total number of days, which is an order of magnitude higher than the computed over all grid boxes. It is also important to consider intraseasonal variability, as the results show that the percentage is even higher when computed for each month separately than for the entire season. For days classified in the NAO- cluster, a maximum value equal to 19.47% in February, but well over 15% from December to March at various locations. It is common to all the regimes that the maximum occurs during the coldest months because the threshold was computed for the entire season. November and March are the warmest months; hence, they register less cold events. During SCAND, high frequencies are located mostly in Northern and Central Europe, but values are lower than for the NAO-. Joint probabilities are high during the AR mainly over Scandinavia and Northwestern Europe. On average, the lowest ratio of events is observed during NAO+, the WR associated with warmer than normal conditions. Nevertheless, at some locations in the south where winds are weak and temperatures are cold, frequencies are relatively high.

It is fairly clear from Figure 4.14 that the geographical distribution of the joint probabilities not only varies among the regimes but also intraseasonal. For instance, during SCAND conditions, the high probabilities of joint events move from southwest in December to northeast in February, covering the largest region in January. In the case of the NAO-, high probabilities are located in Northern Europe and Scandinavia during, extending over a bigger region in January. Only a few locations are affected by high probabilities in the transition months. As can be seen, the probability values and local maximums' locations also present a substantial seasonal

variation.

Overall, Figure 4.14 illustrates that both sets of predictors, the months and the regimes, provide relevant information for modeling the joint probabilities. At the same time, the regions where high probabilities are found are located over the continent or coastal areas, coinciding with the location of wind farms. A high density of wind farms is situated over Northern Europe, mostly over Germany, Denmark, and Britain. These regions are the most affected by joint events during NAO- and SCAND in the coldest months of the year.

Cold events and weak wind events are statistically dependent, meaning that the occurrence of one of them affects the occurrence of the other. Dependence is proven by verifying that they do not satisfy the mathematical definition of independence and by applying the chi-square test for independence, which shows that the events are independent for a 10% level of confidence over most of the domain. Correlation is a measure of how two or more variables are related to one another. It is useful because it can indicate a predictive relationship that can be exploited in practice. In this case, it motivates the study of the joint probabilities of low temperature and low wind speed events by the analysis of Gaussian copula functions. Figures 4.18 and 4.19 show that cold events and weak wind events are positively correlated over land (except some high altitude regions) and at the coasts and shallow waters. Examples of these regions are the North Sea, which is located mostly on the European Continental shelf, and the Baltic Sea. Seasonal variations of correlation values are in general small. In regions where no cold event has been registered, correlations cannot be computed. This frequently happens in November when temperatures are higher than during the other months of the season. The SCAND pattern shows higher probabilities in Britain, France, and the Iberian Peninsula in December and January, whereas the highest correlations over Scandinavia take place in February. During the NAO-, Scandinavia presents high correlations from December to March, but a peak is observed over Finland and Northwestern Europe in December, and over Northern Europe in January. These correlation patterns are reflected in the illustrations of joint probabilities (4.14).

Marginals are modeled with logistic regressions using two sets of binary predictors, one for the months considered in the winter season and one for the WRs. Given that these are the only explanatory variables, the modeled marginals are equal to the observed frequencies. The bss are computed with respect to a seasonality model that only considers the set of monthly predictors. Positive values indicate that large-scale systems impact extreme temperature and wind speed events and, as a consequence, on the peaks of demand and production of energy.

The significance of the scores is assessed with a permutation test, and significant values at the 10% level are exhibited in Figures 4.22 and 4.23. These figures show that considering the WRs improves the performance of the models for marginal probabilities over large geographical areas. In particular, in the regions that we are interested: where the joint probabilities are high.

The high correlations between cold events and weak wind events in regions where marginal probabilities are high motivate the selection of the Gaussian copula framework for modeling joint extreme events. This thesis describes a flexible methodology that can be adapted to different models for marginal distributions. The applicability of the methodology has been tested on the marginals modeled with logistic regression. The BS computed for modeled joint probabilities with respect to the observations is a measure of how well the observed data fit the modeled probabilities. The BSS were computed with respect to a model without correlations. To assess the significance of the results, a permutation test was performed at the 10% level. For the NAO-, the pattern with the highest probabilities, the BSS are, in general, positive in Scandinavia and Northern Europe, with a maximum in January, located over Northern Europe, where correlations also have a local maximum. The scores computed for the SCAND have a noisier signal, tend also to be positive where high probabilities are modeled. In the case of the AR, BSS are positive over major parts of Europe, not only in Scandinavia, where joint events are expected to happen more frequently. The positive values for the BSS computed for a reference model with no correlation between the events indicate that considering the dependency between the events improves the performance. These are promising findings and inspire further research.

It is important to remark that the results presented in this thesis depend strongly on the dataset used. There is an unavoidable degree of uncertainty affecting reanalysis datasets, being the most significant disagreements in DJF encountered within continental areas (Ramon et al., 2019). Reanalyses have good coverage and long records, but the long-term means of wind speeds at hub height derived from ERA5 are biased compared to tall tower observations. Representativeness can also be a problem because wind farms are often located in places where the wind is higher than its surroundings, and thus the mean value of the grid box might be inaccurate. It is expected that the higher resolution of the ERA5 dataset compared to other reanalyses provides more accurate results.

One of the challenges of working with wind speed data from reanalyses is that it is available at 10 meters, whereas typical hub heights are 80-120 meters. To estimate hub height winds from surface winds, an extrapolation method is typically used. The power law, as in Wiel,

Stoop, et al. (2019), was tested for the computation of marginal low wind speed probabilities (not shown), but discarded because it adds even more uncertainty to the wind-speed data, which is already biased. Ramon et al. (2019) discourage the use of global reanalyses to estimate mean winds because of the high uncertainty derived from the comparison with tall tower seasonal averages. In spite of that, they show that ERA5 outperforms all other reanalysis datasets compared in their paper (ERA-Interim, the Japanese 55-year Reanalysis (JRA55), the Modern Era Retrospective Analysis for Research and Applications-2 (MERRA2), and the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Reanalysis 1 (R1)). Future research should investigate the potential effects of using reanalysis data in a copula framework.

The methodology provided can be adapted to different definitions of thresholds. The skill can potentially be improved by modeling the marginal probabilities with more complex algorithms and taking into account more variables. Future studies should also aim to explore how solar energy can complement wind power in situations where probabilities of co-occurrence of cold and weak wind events are high, in order to avoid energy shortfall. Future research could also examine the applicability of this methodology to variables such as energy demand and energy shortfall, modeled as in Wiel, Stoop, et al. (2019) and Hannah C. Bloomfield, David J. Brayshaw, and Charlton-Perez (2020).

Overall, the results confirm that the dependence between the variables plays an important role in modeling compound meteorological events, and evidence that low frequency circulation patterns control parts of the distribution of extreme events is provided. The occurrence of these WRs is preconditioned by the Madden-Julian Oscillation (Cassou, 2008), allowing for its medium-range predictability. This means that the range for predicting the probabilities of occurrence of joint events exceeds the limit of around one week, and the promising skill could be of value for decision-makers in the energy sector.

6 Summary and Conclusions

An increase in the share of renewable energy is crucial to move to less carbon-intensive economies. Wind energy has led the recent growth in renewable-based capacity, and it is expected to continue to be the largest source of renewable energy through to 2030 (Sawyer et al., 2017). Both electricity demand and wind power generation are sensitive to meteorological conditions and their variability at several time scales. To guarantee the supply of energy in a high-renewable system, some critical conditions must be considered. This study provides a versatile methodology for assessing the probability of compound meteorological events that will potentially lead to high electricity demand and low wind energy production. In addition, the co-occurrence of low temperature and low wind speed events over Europe conditioned by large-scale atmospheric circulation patterns is characterized. The atmospheric variability is described in terms of Euro-Atlantic WRs constructed with clustering techniques. The events are studied during the winter season, when the WRs are more persistent Cassou, 2008.

Cold temperatures and weak wind speed conditions are associated with high energy demand and low supply; thus, understanding their joint distribution is decisive for preventing energy shortfall. Instead of focusing on average departures from normal conditions, as most studies do, a baseline for modeling multivariate probabilities is established, and historical probabilities are characterized. These is accomplished by analysing hourly reanalysis data. It is well known that persistent anticyclonic circulation patterns over Europe, such as the observed during NAO- and SCAND, impact surface variables, causing colder than normal temperatures and weak winds. The results in this work show that compound events are most likely to occur during blocking conditions, as previous research indicates, although the occurrence during other WRs should not be underestimated. Frequencies higher than 15% of the days are observed associated with the NAO- from December to February. The NAO+ is characterized by being the circulation pattern with the lowest rate of compound events. The study of the observed frequencies reveals that the monthly decomposition of the events leads to higher probabilities compared to studying the occurrence during the whole winter season. In general, probabilities are highest during NAO- in Northern Europe and Scandinavia from December to March, peaking in February. During SCAND, high probabilities are observed already in December at some locations. The maximum is reached in January when high probabilities are expected over all Europe except Scandinavia.

In both cases, the maximum over land is reached in January, whereas the is mostly affected in February. On the other hand, the AR presents higher probabilities in the Mediterranean and the North Atlantic. Compound events can, however, also be expected over land with relatively high frequency. Overall, November is the month with the lowest observed frequencies, while some events can be expected in March over the sea during NAO- and AR.

Modeling marginal distributions with logistic regressions confirms that the WRs control the occurrence of both low temperature and low wind speed events. The BSS computed with respect to a simple climatology model are significant at the 10% level in regions where marginal probabilities are high, where compound events are more likely to occur. The model scores highest where probabilities are very low, in particular, during NAO+.

High correlations (exceeding $\rho = 0.7$) between temperature and wind speed events are found during all the regimes, motivating the second part of this thesis. A flexible framework for computing the joint probabilities of low temperatures and low wind speed events is described and applied to reanalysis data from ERA5, replicating the observed odds ratios. The methodology consists of computing the Gaussian parametric copulas by estimating the parameters with MCMC (D. Hoff, 2007). Marginal probabilities are computed with a logistic regression consisting of two sets of binary variables, one for the months in the winter season and one for the WRs. The marginal probabilities model can easily be adapted to a broader set of explanatory variables and thresholds to define the events. The BSS computed with respect to a model that considers independent variables, shows promising skill where compound events occur more frequently. This aspect of the research suggested that correlations are essential for modeling the occurrence of compound events.

This thesis concludes by arguing that the occurrence of large-scale WRs conditions the occurrence of joint events over Europe during the wintertime, and that these present a strong seasonality. Cold temperature and weak wind events are correlated, and the dependence between the events is statistically significant. A flexible method to compute the joint distribution by parameterizing the correlations independent from the marginal distributions is documented. It is applied to the reanalysis data, but future research should focus on more complex models for the marginals.

Bibliography

Barnston, Anthony G. and Robert E. Livezey (1987). "Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns". In: *Monthly Weather Review*, 115 (6), pp. 1083–1126. DOI: 10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Bessec, Marie and Julien Fouquau (2008). "The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach". In: *Energy Economics*, 30 (5), pp. 2705–2721. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2008.02.003.

Bloomfield, H C, D J Brayshaw, L C Shaffrey, P J Coker, and H E Thornton (Dec. 2016). "Quantifying the increasing sensitivity of power systems to climate variability". In: *Environmental Research Letters*, 11 (12), p. 124025. DOI: 10.1088/1748-9326/11/12/124025.

— (May 2018). "The changing sensitivity of power systems to meteorological drivers: a case study of Great Britain". In: *Environmental Research Letters*, 13 (5), p. 054028. DOI: 10.1088/1748-9326/aabff9.

Bloomfield, Hannah C., David J. Brayshaw, and Andrew J. Charlton-Perez (2020). "Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types". In: *Meteorological Applications*, 27 (1), e1858. DOI: 10.1002/met.1858.

Brayshaw, David James, Alberto Troccoli, Rachael Fordham, and John Methven (2011). "The impact of large scale atmospheric circulation patterns on wind power generation and its potential predictability: A case study over the UK". In: *Renewable Energy*, 36 (8), pp. 2087–2096. DOI: 10.1016/j.renene.2011.01..

Brier, Glenn W. (1950). "VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY". In: *Monthly Weather Review*, 78 (1), pp. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Cassou, Christophe (2008). "Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation". In: *Nature*, 455, pp. 523–527. DOI: 10.1038/nature07286.

Cassou, Christophe (2010). "Euro-Atlantic regimes and their teleconnections". In: *ECMWF Annual Seminar 2010*.

Collins, Seán, Paul Deane, Brian [Ó Gallachóir], Stefan Pfenninger, and Iain Staffell (2018). "Impacts of Inter-annual Wind and Solar Variations on the European Power System". In: *Joule*, 2 (10), pp. 2076–2090. ISSN: 2542-4351. DOI: https://doi.org/10.1016/j.joule.2018.06.020.

Cortesi, Nicola, Verónica Torralba, Nube Gonzalez-Reviriego, Albert Soret, and Francisco Doblas-Reyes (Apr. 2019). "Characterization of European wind speed variability using weather regimes". In: DOI: 10.13140/RG.2.2.21743.48809.

CPC 2012. *Northern Hemisphere Teleconnection Patterns*. URL: https://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml.

Cradden, Lucy C., Frank McDermott, Laura Zubiate, Conor Sweeney, and Mark O'Malley (2017). "A 34-year simulation of wind generation potential for Ireland and the impact of large-scale atmospheric pressure patterns". In: DOI: 10.1016/j.renene.2016.12.079.

D. Hoff, Peter (June 2007). "Extending the rank likelihood for semiparametric copula estimation". In: *Ann. Appl. Stat.* 1 (1), pp. 265–283. DOI: 10.1214/07-AOAS107. URL: https://doi.org/10.1214/07-AOAS107.

Ely, Caroline R., David J. Brayshaw, John Methven, James Cox, and Oliver Pearce (2013). "Implications of the North Atlantic Oscillation for a UK–Norway Renewable power system". In: *Energy Policy*, 62, pp. 1420–1427. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2013.06.037.

Ferranti, Laura, Susanna Corti, and Martin Janousek (2015). "Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector". In: *Quarterly Journal of the Royal Meteorological Society*, 141 (688), pp. 916–924. DOI: 10.1002/qj.2411. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2411.

Gräler, Dr. Benedikt. *Copulatheque*. URL: https://copulatheque.shinyapps.io/copulas/.

Grams, Christian, Remo Beerli, Stefan Pfenninger, Iain Staffell, and Heini Wernli (July 2017). "Balancing Europe's wind-power output through spatial deployment informed by weather regimes". In: *Nature Climate Change*, 7. DOI: 10.1038/nclimate3338.

Hersbach, Hans, W Bell, P. Berrisford, Andras Horányi, Muñoz-Sabater J., J. Nicolas, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, and Dick Dee (Apr. 2019). "Global reanalysis: goodbye ERA-Interim, hello ERA5". In: (159), pp. 17–24. DOI: 10.21957/vf291hehd7. URL: https://www.ecmwf.int/node/19027.

Hoyer, S. and J. Hamman (2017). "xarray: N-D labeled arrays and datasets in Python". In: *Journal of Open Research Software*, 5 (1). DOI: 10.5334/jors.148. URL: http://doi.org/10.5334/jors.148.

Huber, Matthias, Desislava Dimkova, and Thomas Hamacher (2014). "Integration of wind and solar power in Europe: Assessment of flexibility requirements". In: *Energy*, 69 (C), pp. 236–246. DOI: 10.1016/j.energy.2014.02..

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering*, 9 (3), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

Hurrell, J.W, Yochanan Kushnir, Geir Ottersen, and Martin Visbeck (Jan. 2003). *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*. Vol. 134. DOI: 10.1029/GM134.

Jerez, Sonia and Ricardo M Trigo (Nov. 2013). "Time-scale and extent at which large-scale circulation modes determine the wind and solar potential in the Iberian Peninsula". In: *Environmental Research Letters*, 8 (4), p. 044035. DOI: 10.1088/1748-9326/8/4/044035.

Jones, Eric, Travis Oliphant, Pearu Peterson, et al. (2001–). *SciPy: Open source scientific tools for Python*. URL: http://www.scipy.org/.

Liobikienė, Genovaitė and Mindaugas Butkus (June 2017). "The European Union possibilities to achieve targets of Europe 2020 and Paris agreement climate policy". In: *Renewable Energy*, 106, pp. 298–309. DOI: 10.1016/j.renene.2017.01.036.

Matsueda, Mio and T. N. Palmer (2018). “Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts”. In: *Quarterly Journal of the Royal Meteorological Society*, 144 (713), pp. 1012–1027. DOI: 10.1002/qj.3265.

Met Office (2010 - 2015). *Cartopy: a cartographic python library with a matplotlib interface*. Exeter, Devon. URL: http://scitools.org.uk/cartopy.

Michelangeli, Paul-Antoine, Robert Vautard, and Bernard Legras (1995). “Weather Regimes: Recurrence and Quasi Stationarity”. In: *Journal of the Atmospheric Sciences*, 52 (8), pp. 1237–1256. DOI: 10.1175/1520-0469(1995)052<1237:WRRAQS>2.0.CO;2.

Neal, Robert, David Fereday, Ric Crocker, and Ruth E. Comer (2016). “A flexible approach to defining weather patterns and their application in weather forecasting over Europe”. In: *Meteorological Applications*, 23 (3), pp. 389–400. DOI: 10.1002/met.1563.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Pinson, Pierre (Nov. 2013). “Wind Energy: Forecasting Challenges for Its Operational Management”. In: *Statist. Sci.* 28 (4), pp. 564–585. DOI: 10.1214/13-STS445.

Quadrelli, Roberta and John Wallace (Oct. 2004). “A Simplified Linear Framework for Interpreting Patterns of Northern Hemisphere Wintertime Climate Variability”. In: *Journal of Climate - J CLIMATE*, 17, pp. 3728–3744. DOI: 10.1175/1520-0442(2004)017<3728:ASLFFI>2.0.CO;2.

Ramon, Jaume, Llorenç Lledó, Verónica Torralba, Albert Soret, and Francisco J. Doblas-Reyes (2019). “What global reanalysis best represents near-surface winds?” In: *Quarterly Journal of the Royal Meteorological Society*, 145 (724), pp. 3236–3251. DOI: 10.1002/qj.3616. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3616. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3616.

Ravestein, P., G. [van der Schrier], R. Haarsma, R. Scheele, and M. [van den Broek] (2018). "Vulnerability of European intermittent renewable energy supply to climate change and climate variability". In: *Renewable and Sustainable Energy Reviews*, 97, pp. 497–508. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2018.08.057.

Reinhold, Brian B. and Raymond T. Pierrehumbert (1982). "Dynamics of Weather Regimes: Quasi-Stationary Waves and Blocking". In: *Monthly Weather Review*, 110 (9), pp. 1105–1145. DOI: 10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2.

Rogelj, Joeri, Gunnar Luderer, Robert C. Pietzcker, Elmar Kriegler, Michiel Schaeffer, Volker Krey, and Keywan Riahi (June 2015). "Energy system transformations for limiting end-of-century warming to below 1.5 C". In: *Nature Climate Change*, 5 (10), pp. 519–527. DOI: 10.1038/nclimate2572.

Sawyer, Steve, lauha fried lauha, shruti shukla shruti, and Qiao Liming (Apr. 2017). *Global Wind Report 2016 - Annual Market Update*.

Sinden, Graham (Jan. 2007). "Characteristics of the UK wind resource: Long-term patterns and relationship to electricity demand". In: *Energy Policy*, 35, pp. 112–127. DOI: 10.1016/j.enpol.2005.10.003.

Sklar, M. (1959). "Fonctions de répartition à *n* dimensions et leurs marges". In: *Publ. Inst. Statist. Univ. Paris*, 8, pp. 229–231.

Straus, David M., Susanna Corti, and Franco Molteni (2007). "Circulation Regimes: Chaotic Variability versus SST-Forced Predictability". In: *Journal of Climate*, 20 (10), pp. 2251–2272. DOI: 10.1175/JCLI4070.1.

Taylor, James and R. Buizza (Mar. 2003). "Using weather ensemble predictions in electricity demand forecasting". In: *International Journal of Forecasting*, 19, pp. 57–70. DOI: 10.1016/S0169-2070(01)00123-6.

team, The pandas development (Feb. 2020). *pandas-dev/pandas: Pandas*. Version latest. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

Thompson, David W. J. and John M. Wallace (1998). "The Arctic oscillation signature in the wintertime geopotential height and temperature fields". In: *Geophysical Research Letters*, 25 (9), pp. 1297–1300. DOI: 10.1029/98GL00950. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/98GL00950. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98GL00950.

Thornton, Hazel, Adam Scaife, Brian Hoskins, and David Brayshaw (June 2017). "The relationship between wind power, electricity demand and winter weather patterns in Great Britain". In: *Environmental Research Letters*, 12, p. 064017. DOI: 10.1088/1748-9326/aa69c6.

Trenberth, Kevin E., Grant W. Branstator, David Karoly, Arun Kumar, Ngar-Cheung Lau, and Chester Ropelewski (1998). "Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures". In: *Journal of Geophysical Research: Oceans*, 103 (C7), pp. 14291–14324. DOI: 10.1029/97JC01444. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/97JC01444.

Van Der Walt, Stefan, S Chris Colbert, and Gael Varoquaux (2011). "The NumPy array: a structure for efficient numerical computation". In: *Computing in Science & Engineering*, 13 (2), p. 22.

Vautard, R. (1990). "Multiple Weather Regimes over the North Atlantic: Analysis of Precursors and Successors". In: *Monthly Weather Review*, 118 (10), pp. 2056–2081. DOI: 10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2.

Wallace, John M. and David S. Gutzler (1981). "Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter". In: *Monthly Weather Review*, 109 (4), pp. 784–812. DOI: 10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wheeler, Matthew C. and Harry H. Hendon (2004). "An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction". In: *Monthly Weather Review*, 132 (8), pp. 1917–1932. DOI: 10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wiel, Karin van der, Hannah C Bloomfield, Robert W Lee, Laurens P Stoop, Russell Blackport, James A Screen, and Frank M Selten (Sept. 2019). "The influence of weather regimes on

European renewable energy production and demand". In: *Environmental Research Letters*, 14 (9), p. 094010. DOI: 10.1088/1748-9326/ab38d3.

Wiel, Karin van der, Laurens Stoop, Bas van Zuijlen, R Blackport, M Broek, and F Selten (May 2019). "Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall". In: *Renewable and Sustainable Energy Reviews*, 111, pp. 261–275. DOI: 10.1016/j.rser.2019.04.065.

Wilks, D.S. (2006). *Statistical Methods in the Atmospheric Sciences*. International Geophysics. Elsevier Science. ISBN: 9780127519661.

Wohland, J., M. Reyers, J. Weber, and D. Witthaut (2017). "More homogeneous wind conditions under strong climate change decrease the potential for inter-state balancing of electricity in Europe". In: *Earth System Dynamics*, 8 (4), pp. 1047–1060. DOI: 10.5194/esd-8-1047-2017.

Zeyringer, Marianne, James Price, Birgit Fais, Pei-Hao Li, and Ed Sharp (May 2018). "Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather". In: *Nature Energy*, 3. DOI: 10.1038/s41560-018-0128-x.

Zubiate, Laura, Frank McDermott, Conor Sweeney, and Mark O'Malley (2017). "Spatial variability in winter NAO–wind speed relationships in western Europe linked to concomitant states of the East Atlantic and Scandinavian patterns". In: *Quarterly Journal of the Royal Meteorological Society*, 143 (702), pp. 552–562. DOI: 10.1002/qj.2943.

8 Appendix

A Challenges of working with big datasets

The framework explored in this thesis relies on the availability of high resolution-data, but working with big datasets imposes challenges related to storing and processing it. Given that the spatial resolution of the ERA5 data set is $0.25^\circ \times 0.25^\circ$ and that the domain selected covers the region defined by $-35 \circ N - 72^\circ N, 11^\circ W - 40^\circ E$, the grid on which the results were accomplished has size 149×205, this is, 30545 locations. On the time dimension, 47 years of hourly data are processed and aggregated to daily level. This can lead to space and time efficiency problems since learning methods are computation- and memory-intensive.

Some particular issues addressed during this study are:

- Data integration: since the data analyzed comes from different sources, i.e., the ERA5 reanalysis data and the WRs provided. Data formats are different and required different libraries for the analysis.

- Resampling: permutation tests are conceptually simple but computationally expensive, and storing the results is related to upscaling troubles.

- Picking the right formats and libraries: the experiences obtained during this work point toward that R is fast and allows to parallelize operations at a high level. When it comes to Python, xarray is adequate for processing NetCDF data (which is a standard format for geophysical data), and scikit-learn and scipy are useful tools for statistical analysis of data. The problem with these libraries is that they do not integrate well, and conversion to numpy arrays is needed. Pandas DataFrame were tested, but indexing operations are too slow, and the performance is not acceptable for the purpose of this thesis.

B Software

The Python code used to generate the results in this thesis is available in this Github repository: paulina-t/msc-thesis. The code runs on Python version 3.6.7.

B.1 Python packages

The Python package selected for working with the NetCDF is xarray, an open-source project that makes working with labelled multi-dimensional arrays simple. It works on top of NumPy-like arrays and integrates with dask to support parallel computations and streaming computation on datasets that do not fit into memory. Dask divides arrays into many small pieces, called chunks, that fit into memory. This is necessary given the size of the ERA5 files containing the reanalysis data in the Euro-Atlantic domain. Unlike NumPy, which has an eager evaluation, operations on Dask arrays are lazy. The actual computation is controlled by a multi-processing or thread pool, allowing Dask to take full advantage of multiple processors available. In our case, the code runs on 32 CPUs. Statistic analysis is performed with the package Scikit-Learn, which is built upon SciPy. The Scikit-Learn library provides a set of standard algorithms and efficient tools for machine learning and statistical modeling.

Table B.1 lists the dependencies required to run the code.

B.2 R packages

Associations between variables were computing using the R package *Semiparametric Bayesian Gaussian Copula Estimation and Imputation*, sbcop, version 0.980 (D. Hoff (2007). The code runs on R version 3.6.0 (2019-04-26).

The function sbgcop.mcmc estimates semiparametrically the parameters of the Gaussian copula. It takes as an imput an $n \times p$ matrix (missing values are allowed), and returns an array of size $p \times p \times nsamp$, consisting of posterior samples of the correlation matrix.

Table B.1: Python dependencies.

Package	**Version**	**Reference**
numpy	1.16.2	Van Der Walt, Colbert, and Varoquaux, 2011
pandas	0.23.4	team, 2020
matplotlib	2.2.3	Hunter, 2007
cartopy	0.17.0	Met Office, 2010 - 2015
xarray	0.10.9	Hoyer and Hamman, 2017
scikit-learn	0.20.3	Pedregosa et al., 2011
scipy	1.2.1	Jones, Oliphant, Peterson, et al., 2001–
patsy	0.5.1	

C Complementary results

All the figures and tables provided in this thesis can be generated with the notebooks in the GitHub repository.

C.1 Logistic Regression

C.2 Coefficients

The coefficients used to compute the predict the probabilities with the logistic regression models are illustrated in this section.

Climatology model

Temperature

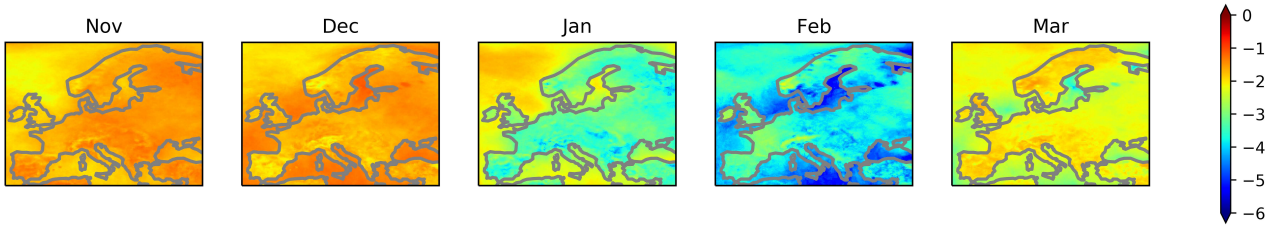

Figure C.1: Seasonality of coefficients of modeled minimum daily surface temperature below the 10^{th} percentile. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = 0$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Temperature

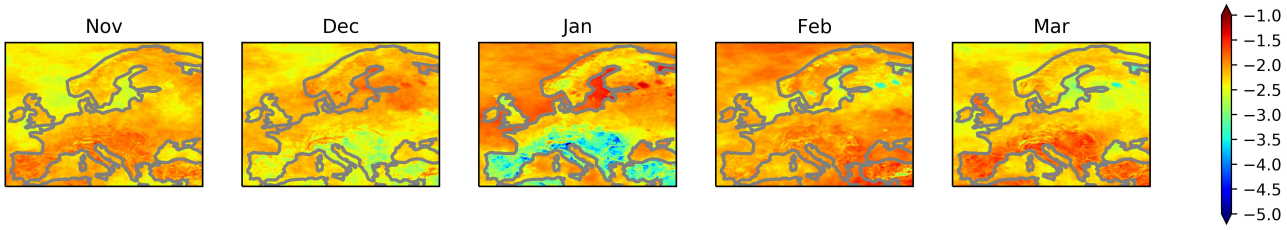

Figure C.2: As C.1 but for maximum daily wind speed at 10 meters. Colours show coefficients, β, and the colorbar is truncated at $\beta = -0.1$ and $\beta = -0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Wind speed

Regime model

Temperature

Temperature C.9

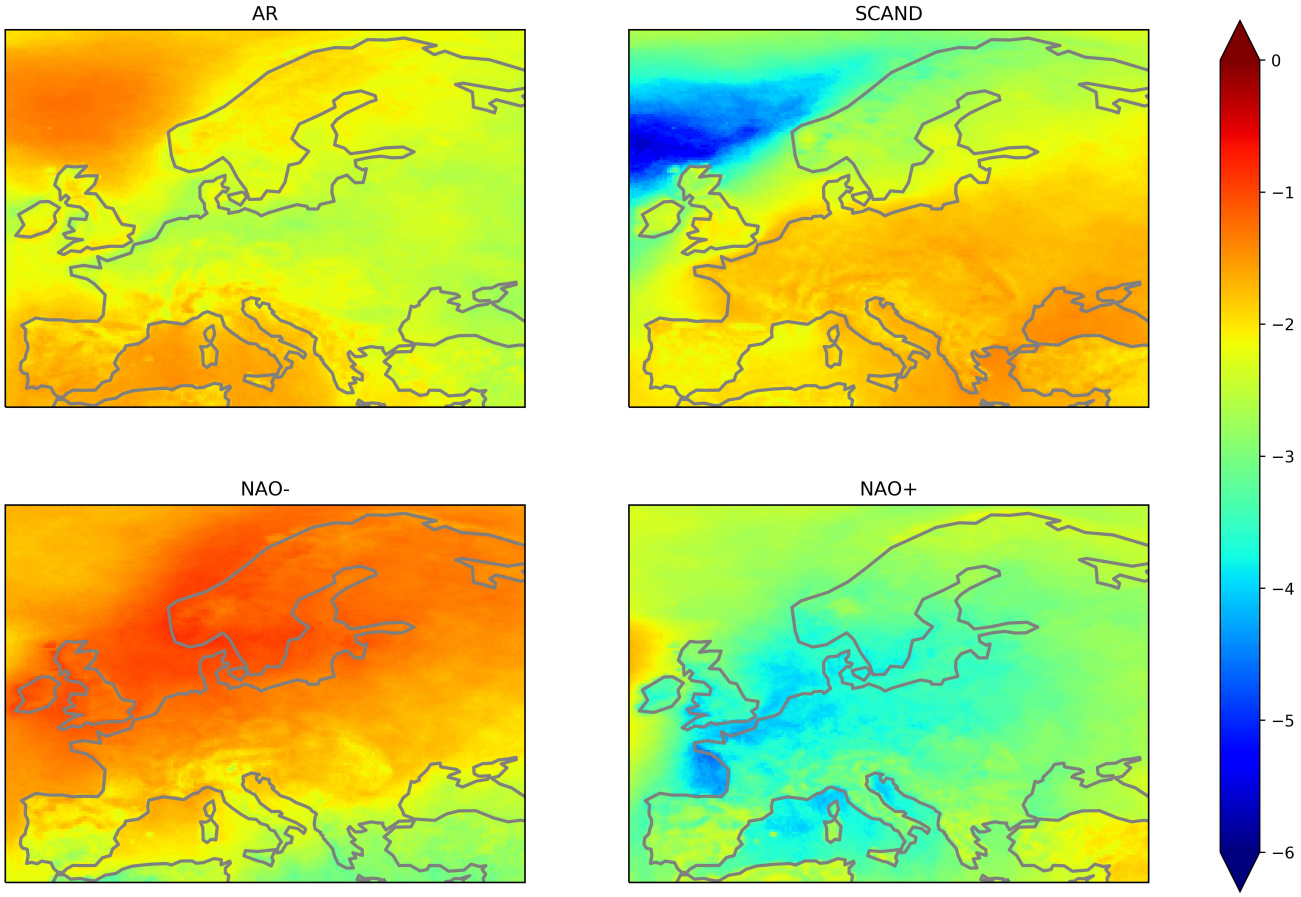

Figure C.3: Coefficients of modeled minimum daily surface temperature below the 10th percentile for each WR. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = 0$ and $\beta = 0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Wind speed

Interaction model

Temperature

Wind speed

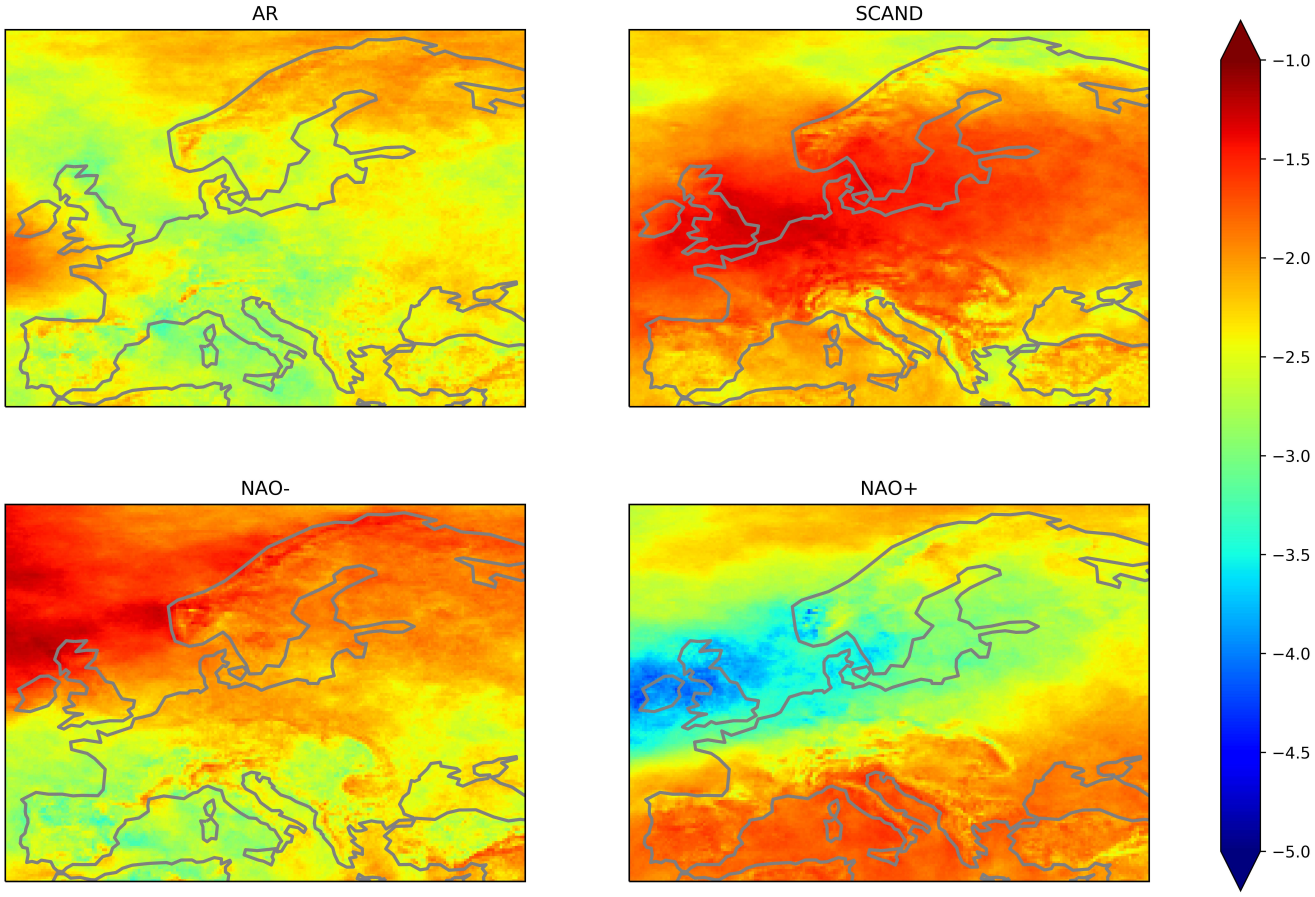

Figure C.4: As C.3 but for maximum daily wind speed at 10 meters. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = -0.1$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

C.2.1 Probabilities

Climatology model

The climatology model consists of a set of dummy variables for each month of the winter season. The monthly probabilities computed with this model are illustrated in Figures C.7 for the temperature variable and C.8 the wind speed variable. The results are very similar to the observed frequencies because the only predictors employed is a set of binary variables for the months of the extended winter season.

Temperature

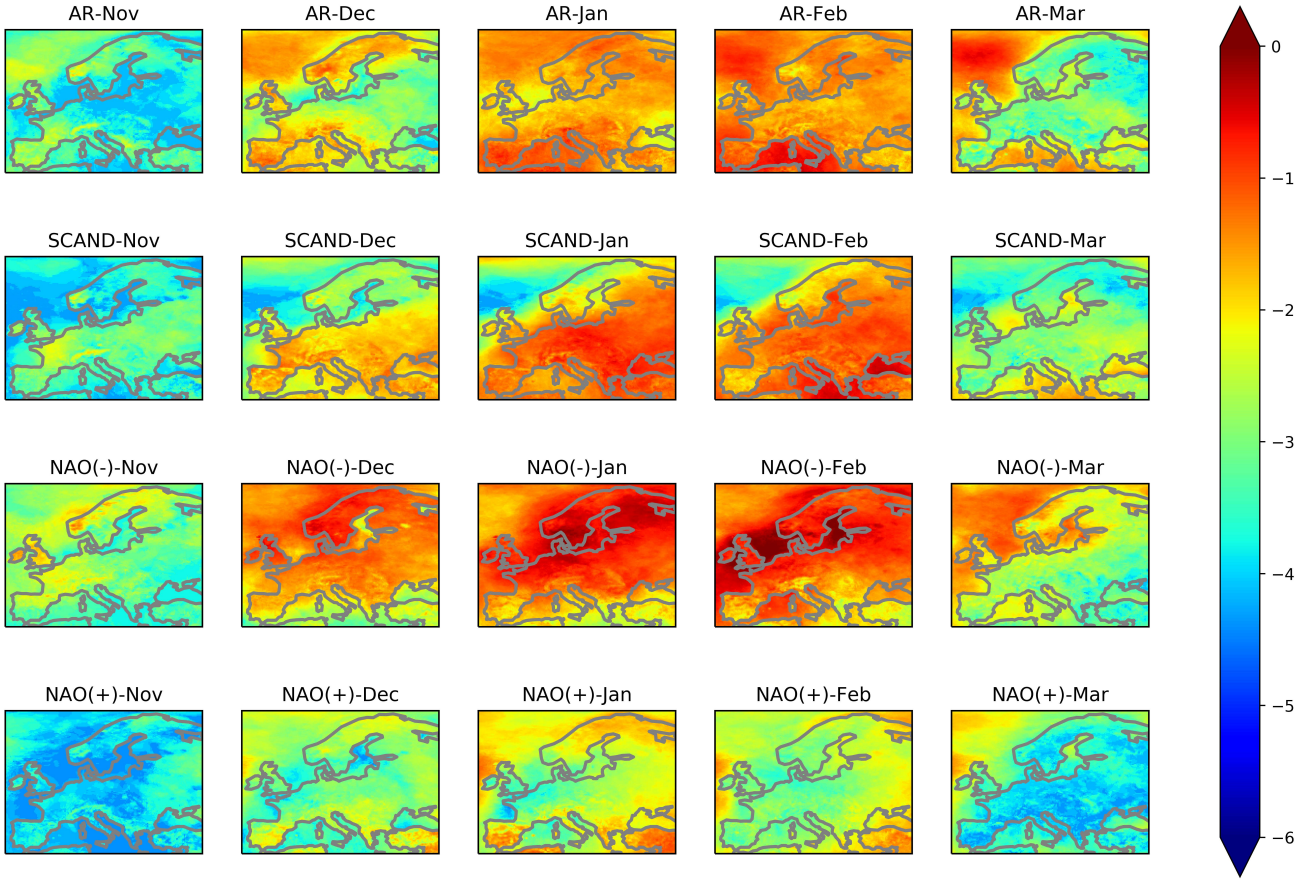

Figure C.5: Coefficients of modeled minimum daily surface temperature below the 10^{th} *percentile for each month and WRs. Colours show coefficients,* β *from the logistic regression model; the color bar is truncated at* $\beta = 0$ *and* $\beta = 0.6$*. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

Wind speed

Regimes model

Modeled probabilities for each regime were computed using only a set of dummy variables for the four WRs. These results are also similar to the ones derived from the empirical distributions, showing a clear preference for the NAO- in Northern Europe and Scandinavia in the case of temperature (Figure 4.8) and high probabilities of low wind associated with anticyclonic circulation in Northwestern Europe during NAO- and Northern and Central Europe during SCAND (4.9).

Temperature C.9

Wind speed

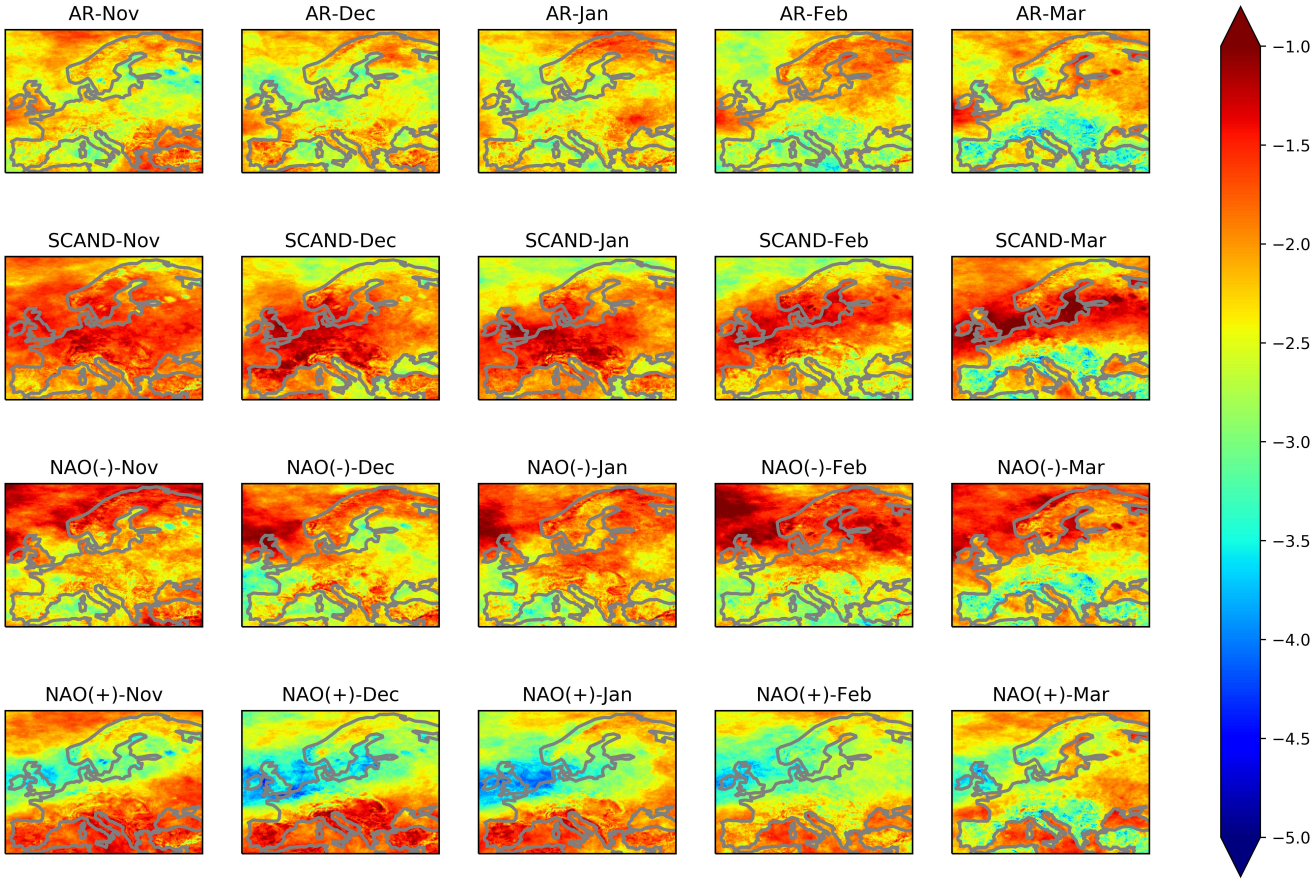

Figure C.6: As C.5 but for maximum daily wind speed at 10 meters. Colours show coefficients, β from the logistic regression model; the color bar is truncated at $\beta = -0.1$ and $\beta = -0.6$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

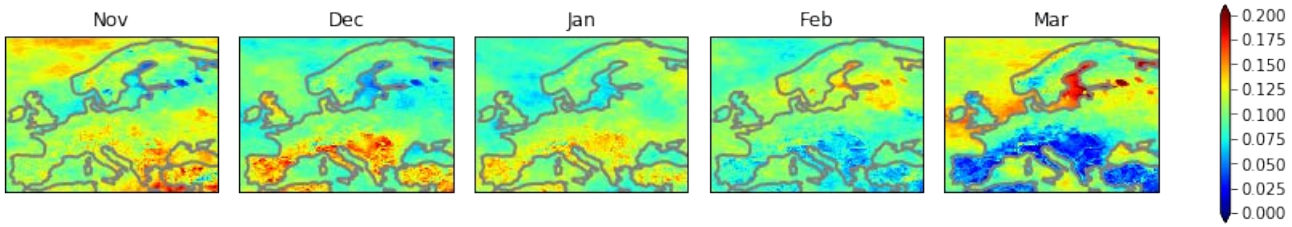

Figure C.7: Seasonality of modeled probabilities of occurrence of minimum daily surface temperature below the 10^{th} percentile for each month. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Interaction model

The seasonality of the marginal probabilities for low temperature events(C.11) and low wind speed (C.12) events was modeled for each regime. Again, the results are similar to the observed

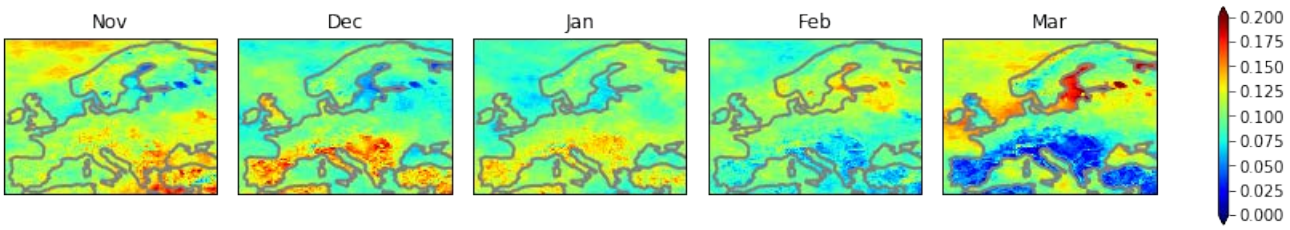

Figure C.8: As C.7 but for maximum daily wind speed at 10 meters. Colours show probabilities, p, and the colorbar is truncated at $p = 0.35$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

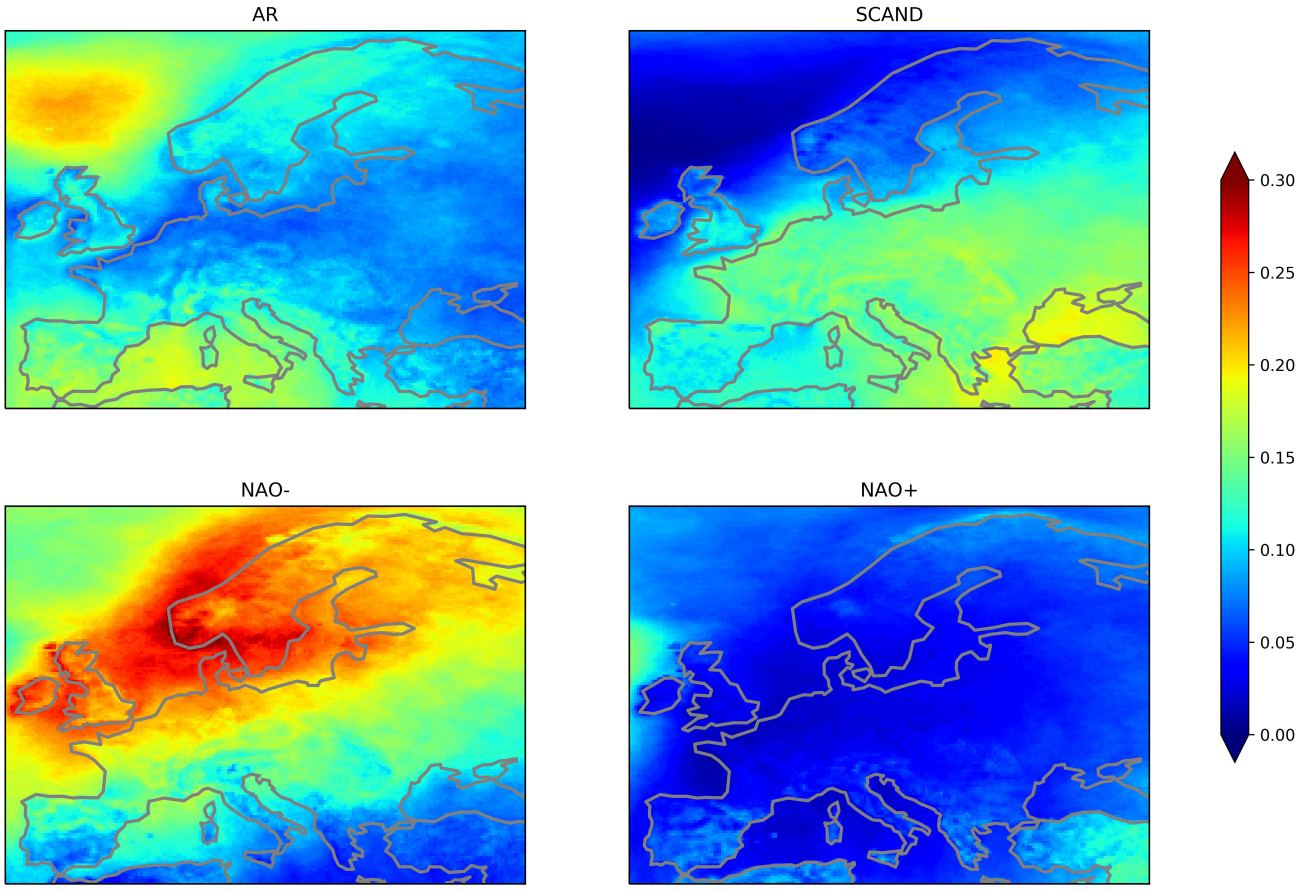

Figure C.9: Modeled probabilities of occurrence of minimum daily surface temperature below the 10th percentile for each WR. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

frequencies.

Temperature Illustration C.11 and Table C.1 show that the model estimates low temperatures (lower than the 10th percentile) at least 20% of the days over large continental areas from

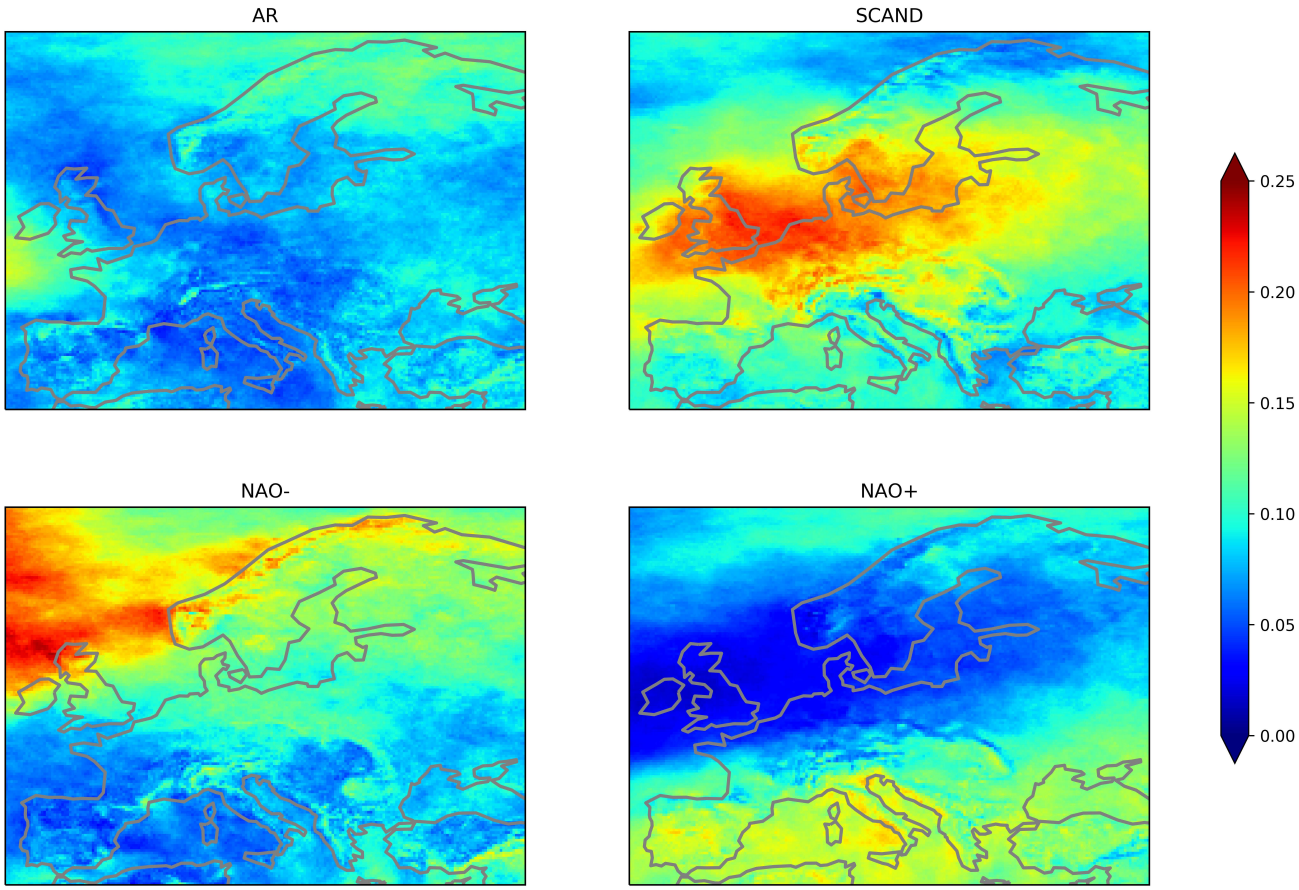

Figure C.10: As C.9 but for maximum daily wind speed at 10 meters. Colours show probabilities, p, and the colorbar is truncated at $p = 0.35$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Table C.1: Maximum probabilities of low temperature events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.1385	0.1321	0.2363	0.0824
Dec	0.2723	0.2661	0.4194	0.2530
Jan	0.3669	0.3851	0.5275	0.2971
Feb	0.4089	0.4460	0.5773	0.2547
Mar	0.3766	0.2012	0.3037	0.2017

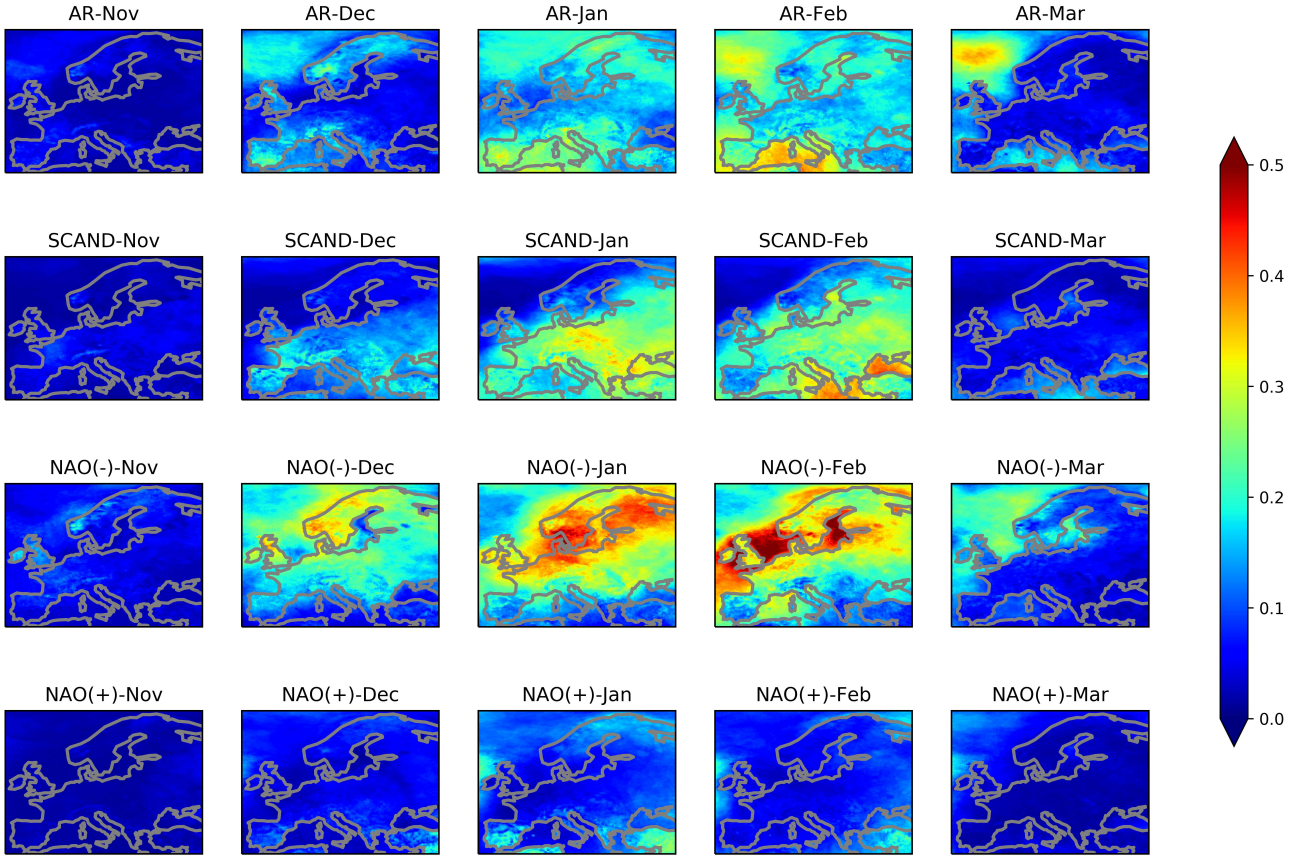

Figure C.11: Modeled probabilities of occurrence of minimum daily surface temperature below the 10^{th} percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

November to February under NAO-, SCAND, and AR conditions. NAO- is the WR with the strongest temperature signal over the continent, followed by the SCAND and the AR. Probabilities are highest during NAO-, increasing from December to February, reaching values as high as $p_{X_{(Feb,NAO-)}} = 0.5773$ in Northern Europe, and a spatial average equal to $p_{\mu_{(Feb,NAO-)}} = 0.2675$ (on average, more than 1/4 of the days are expected to be extremely cold). Values are particularly high over large water bodies such as the North Sea, The Baltic Sea, and the Bay of Biscay. In contrast, in January, the probabilities are in general higher over land and coastal zones than over the open ocean in high latitudes. In the case of the SCAND, large continental areas are expected to experience low temperatures in January and February. The first cold events of the season appear in December, in lower latitudes, over land. The highest probabilities over land are observed in January, with a maximum equal to $p_{X_{(Jan,NAO-)}} = 0.3851$ in Eastern Europe, but over the sea, the maximum is reached in February and equals $p_{X_{(Jan,NAO-)}} = 0.3851$ and the highest

spatial average among all the months equals $p_{\mu_{(Feb,NAO-)}} = 0.4460$ in the Eastern Mediterranean Sea and the Black Sea. This model predicts low probabilities over Scandinavia during the entire season for NAO- days. An important difference between the NAO- and the SCAND, besides the magnitude, is that during NAO- high probabilities are located in high latitudes and low probabilities are located in low latitudes, whereas the opposite takes place during SCAND. The AR presents a zonal band of low probabilities extending from 45N to 60N approximately, with higher probabilities outside this band. A center of relatively high probabilities is situated over the Norwegian Sea during AR in February, intensifying in March. In general, the model predicts low probabilities in November and March, and, in the case of NAO+, throughout the entire winter season.

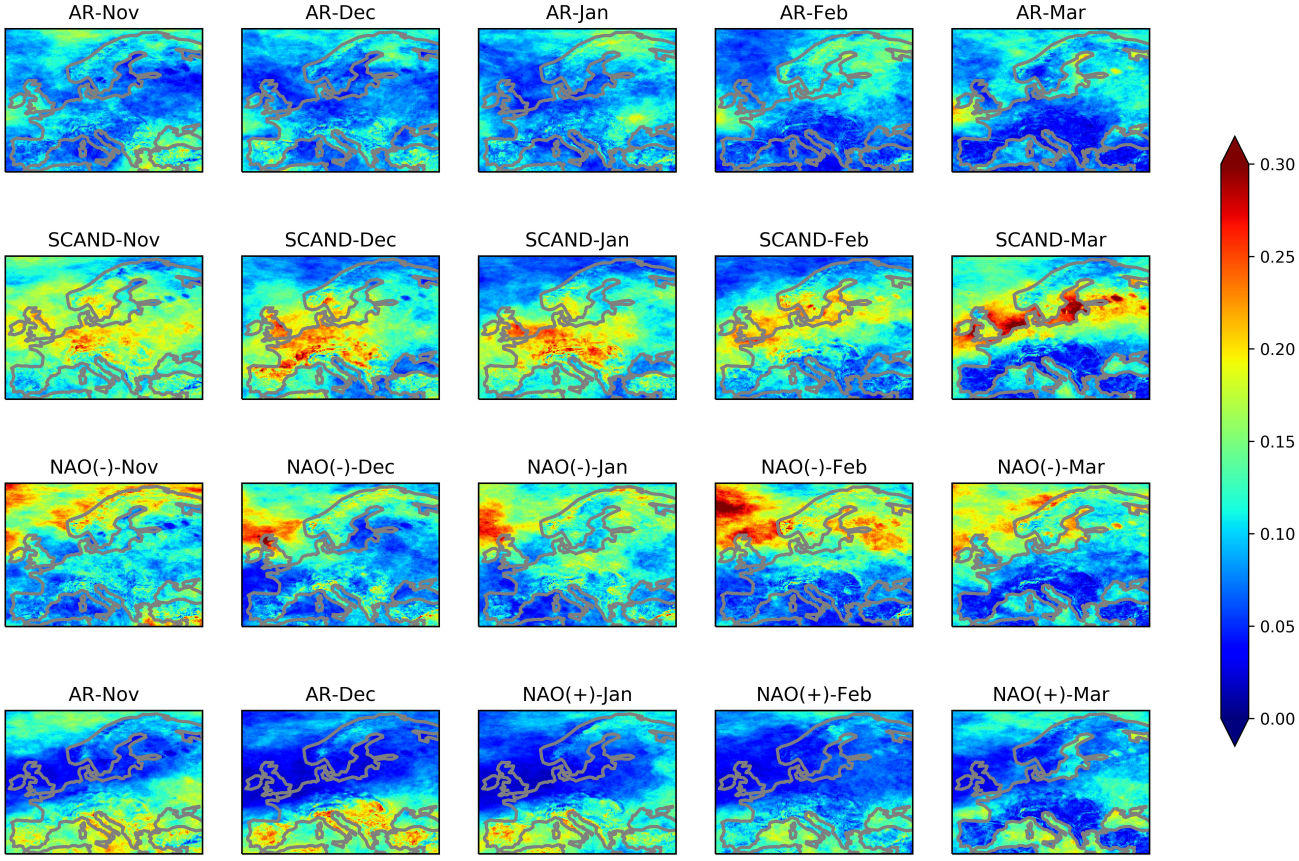

Figure C.12: As C.11 but for maximum daily 10m wind speeds. Colours show probabilities, p, and the color bar is truncated at $p = 0.30$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Wind speed Probabilities of occurrence of low maximum daily wind speed for each month and WR are shown in Figure C.12. In general terms, probabilities are higher during NAO-

Table C.2: Maximum probabilities of low wind speed events (p_X). Area of study: 35N-72N, 11W-40E. Table based on ERA5 data (NDJFM, 1979–2017).

	AR	**SCAND**	**NAO-**	**NAO+**
Nov	0.2264	0.3085	0.3106	0.2800
Dec	0.2307	0.3513	0.3509	0.3112
Jan	0.2147	0.2983	0.2984	0.2794
Feb	0.2032	0.2951	0.3558	0.2139
Mar	0.2431	0.3427	0.2804	0.2249

and SCAND compared to AR and NAO+, and seasonal variations are small. The model shows that both the NAO- and the SCAND have, on average, probabilities between 0.1 and 0.15, but in large regions, 20% of the days are expected to have week winds. The SCAND presents the highest probabilities over the continent, reaching values as high as $p_{X,Dec,SCAND} = 0.3513$. Furthermore, the signal is noisier and seems to depend more on the topography than in the case of low temperature events. The Arctic and Southern Europe are affected by the lowest probabilities during SCAND. The model predicts more extreme wind speed events over land, central and northern Europe for the SCAND in December and January. Contrarily, in March, the model predicts that more days with low wind speeds occur over the ocean in high latitudes and very low probabilities in Southern Europe. The NAO- exhibits a meridional pattern with lower (less than 0.1) probabilities in Central and Southern Europe, and higher probabilities (greater than 0.1) in Northern Europe. Seasonal changes during NAO- are small, although probabilities in the North Atlantic are high from November to February, and decreasing in March. The NAO+ exhibits a weaker and opposite pattern, with higher probabilities in the south and lower probabilities in the north. This pattern is observed from November to February. In March, low probabilities (below 0.1) cover the European continent almost in its totality. The probability field presents a maximum in December in Southern Europe. During AR, from November to March, the model predicts less than 10% of the days with maximum daily wind speeds below the 10th percentile in most parts of the domain. Northern Scandinavia shows higher probabilities

in January and February.

C.3 Comparison of thresholds

C.3.1 Probabilities

As mentioned in the document, the methodology is independent of the arbitrary threshold that defines the events. Here, the marginal probabilities for two different thresholds, the 15^{th} and the 5^{th} percentiles are shown.

Interaction model

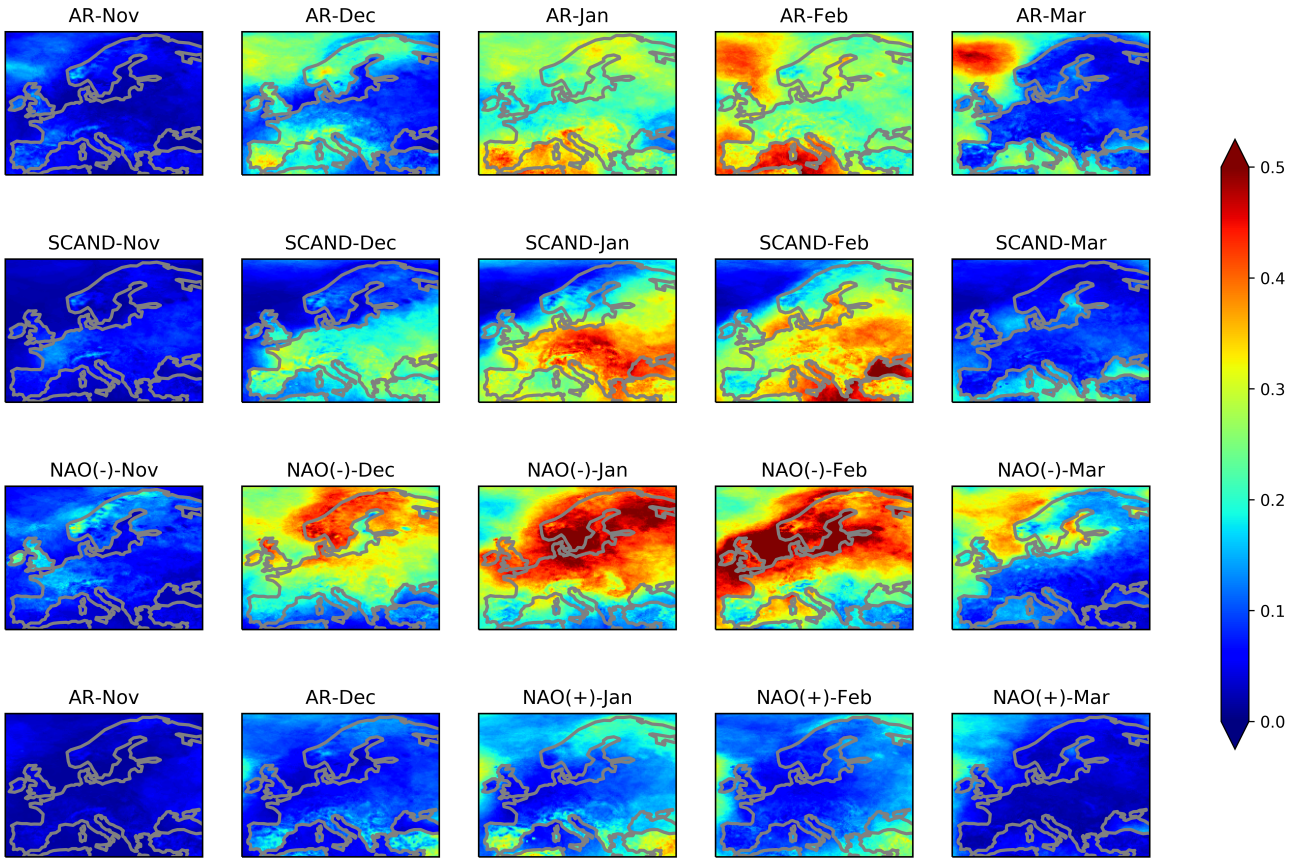

Figure C.13: Modeled probabilities of occurrence of minimum daily surface temperature below the 15th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

Temperature

Wind speed

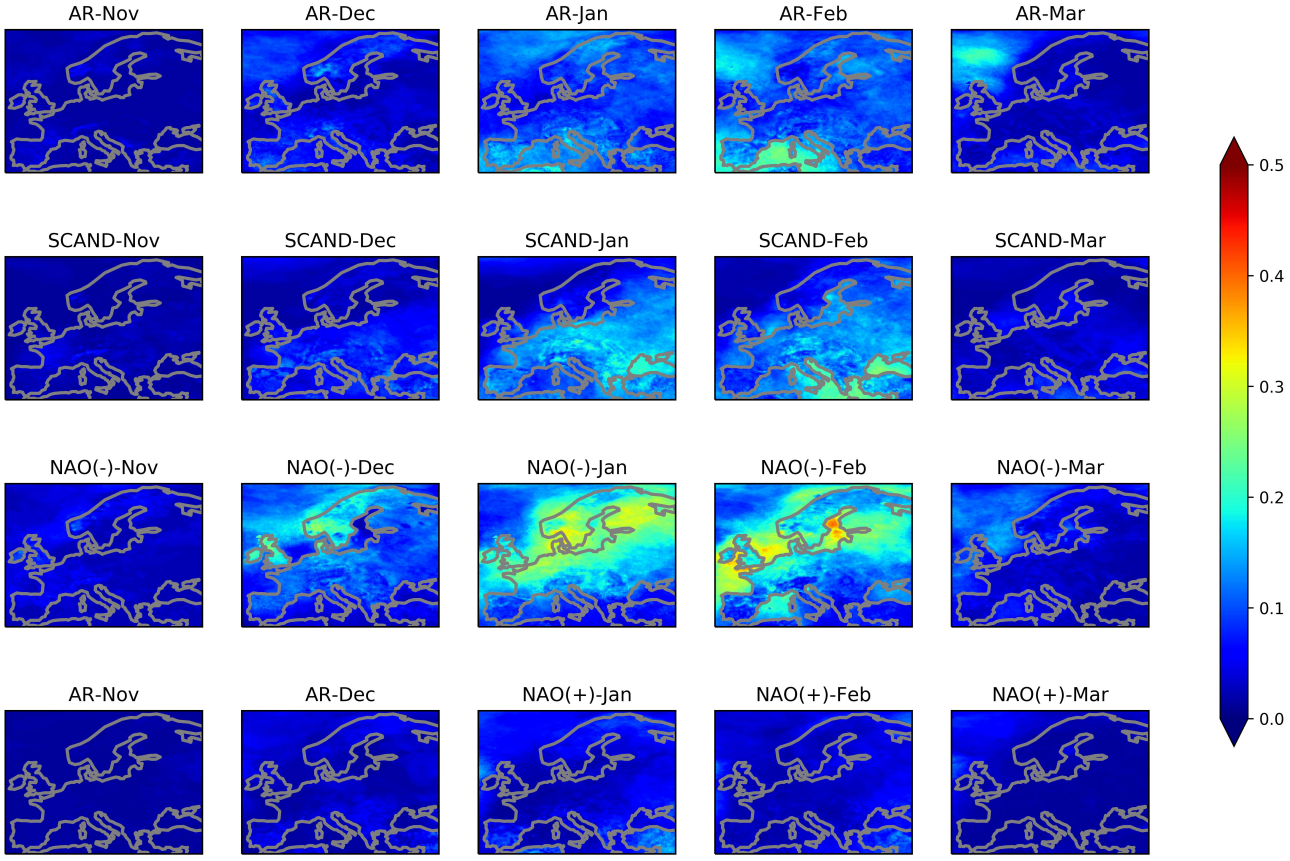

Figure C.14: Modeled probabilities of occurrence of minimum daily surface temperature below the 5th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

C.4 Independence

Independence of cold and weak wind events was quantified by running a chi-square test on each grid box for each regime, for the entire winter season and for each month separately.

C.5 Copula

The joint probabilities calculated for each regime on the entire winter season are replicated using the Gaussian copula approach.

Figure C.21 is generated by computing the copula function from the empirical marginal probabilities for each of the WRs. Coherent with the illustrations of the frequencies of observed events, Figure C.21 shows that NAO- favours the co-occurrence of low temperatures and low wind speeds in the winter season (Nov-Mar) in Northern Europe and Scandinavia, contrasts with the NAO+ that does not show any signal over the domain. During SCAND conditions,

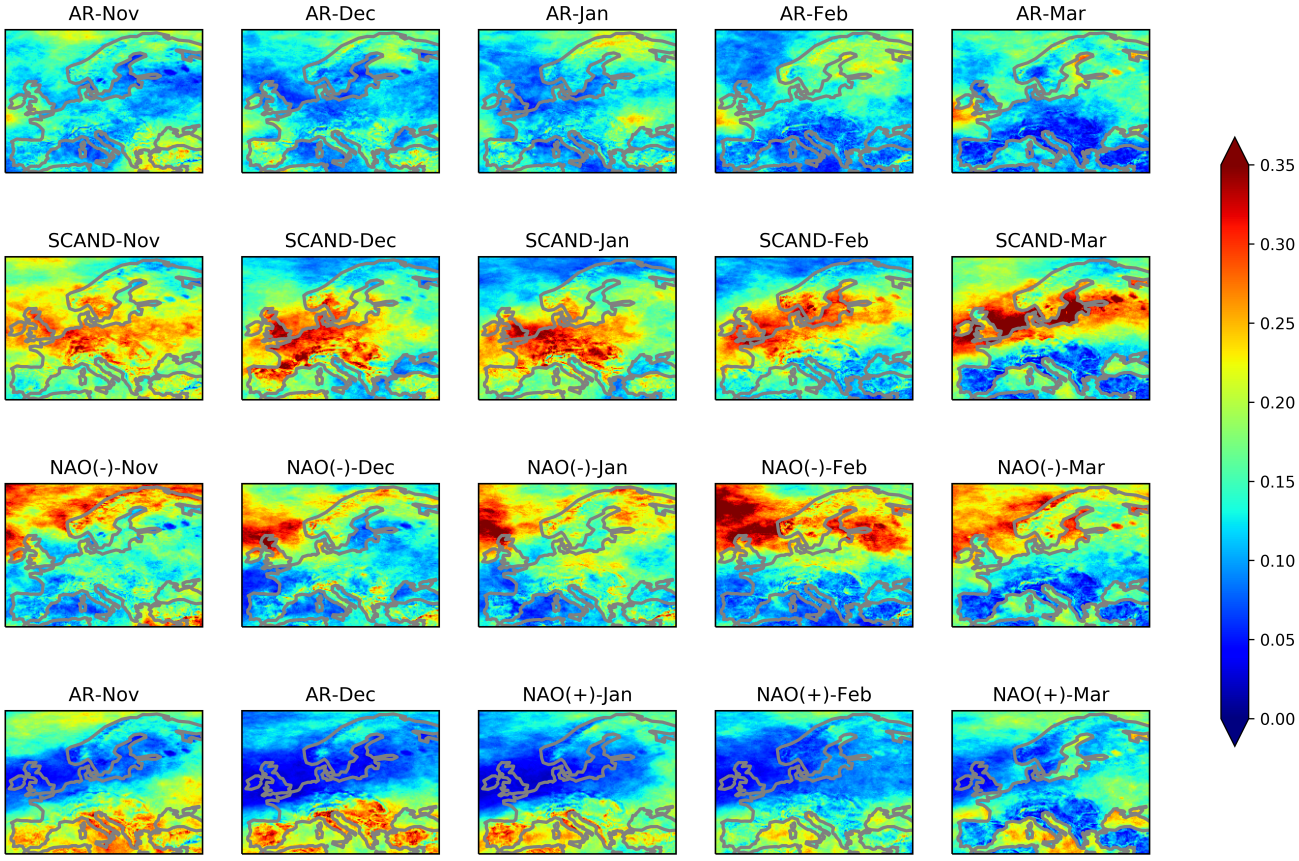

Figure C.15: Modeled probabilities of occurrence of maximum daily 10m wind speeds below the 15th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. *Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

probabilities are higher in Central and Northern Europe, in particular, in France and Britain. The AR pattern presents some signal in Northern Scandinavia and Ireland.

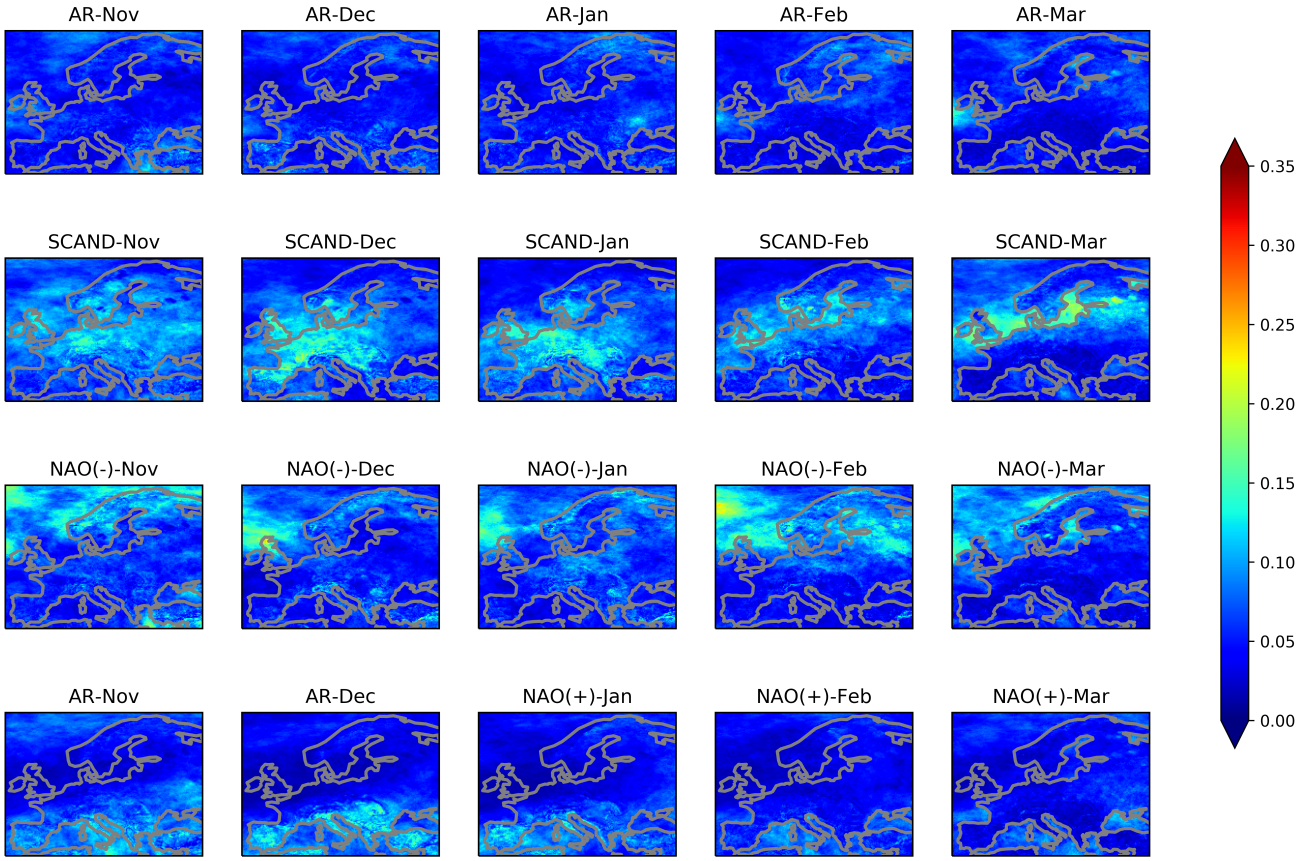

Figure C.16: Modeled probabilities of occurrence of maximum daily 10m wind speeds below the 5th percentile for each month and WRs. Colours show probabilities, p, from the logistic regression model; the color bar is truncated at $p = 0.5$. *Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).*

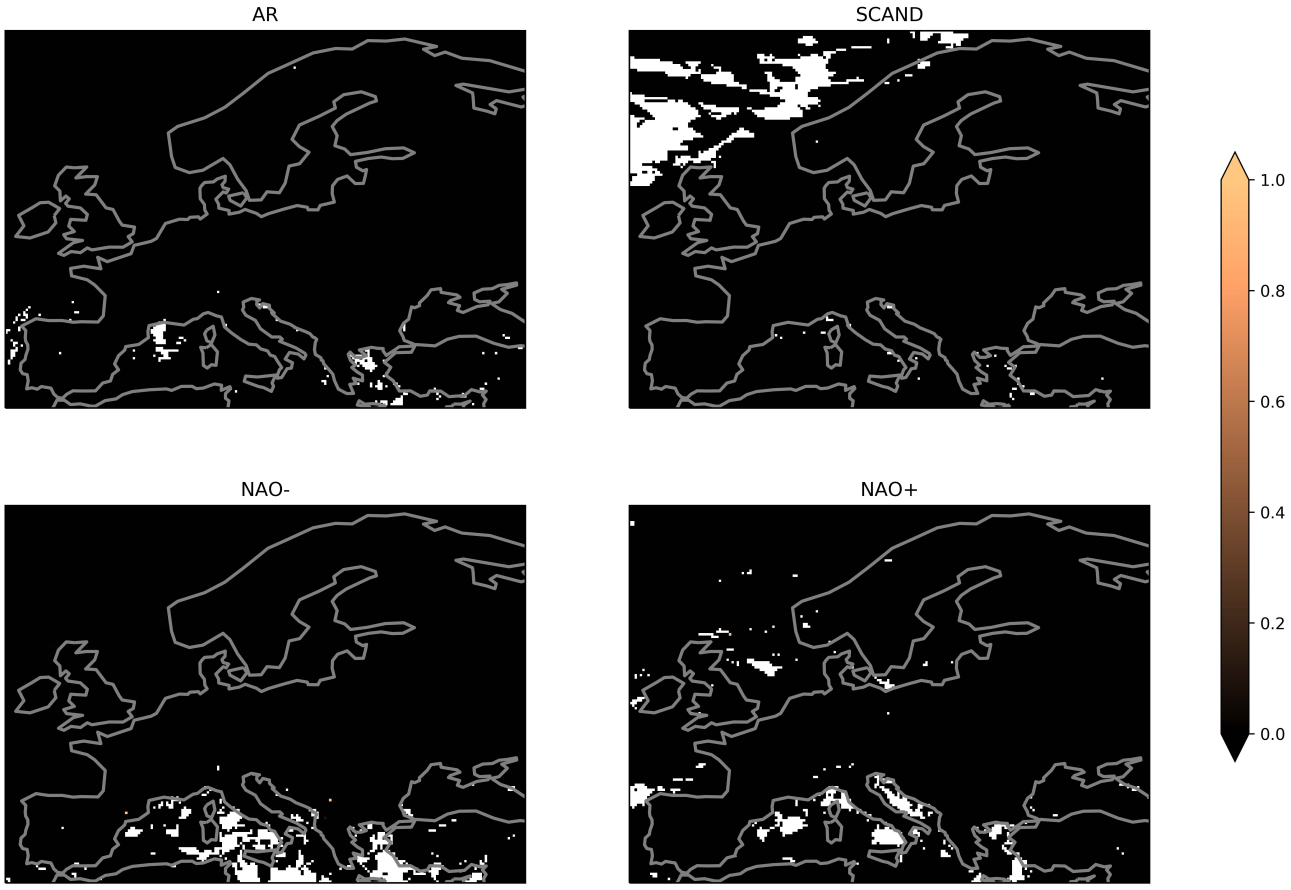

Figure C.17

Figure C.18: Independence between low temperature an low wind speed events by regime. Colours show p-values obtained with the chi-square test. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

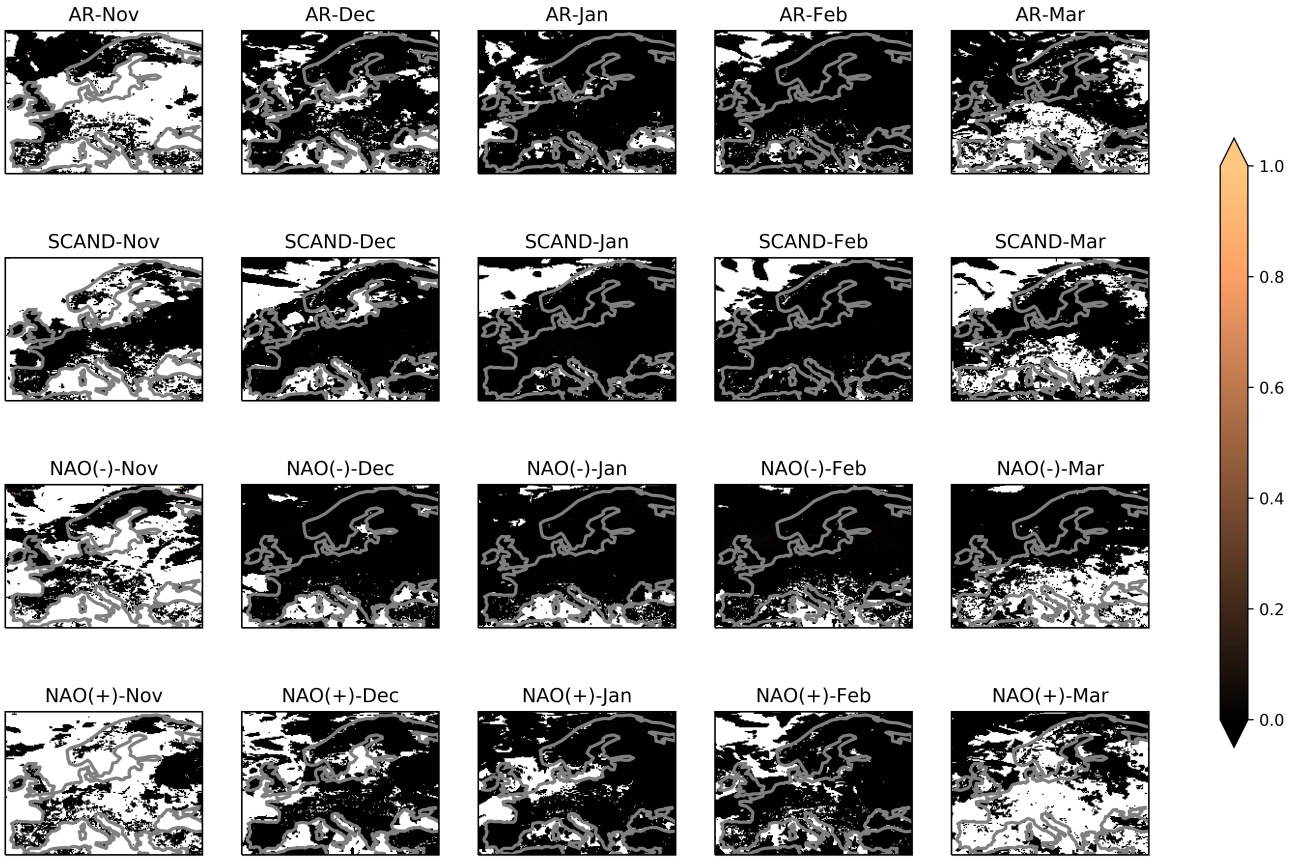

Figure C.19

Figure C.20: Independence between low temperature an low wind speed events by regime and month. Colours show p-values obtained with the chi-square test. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).

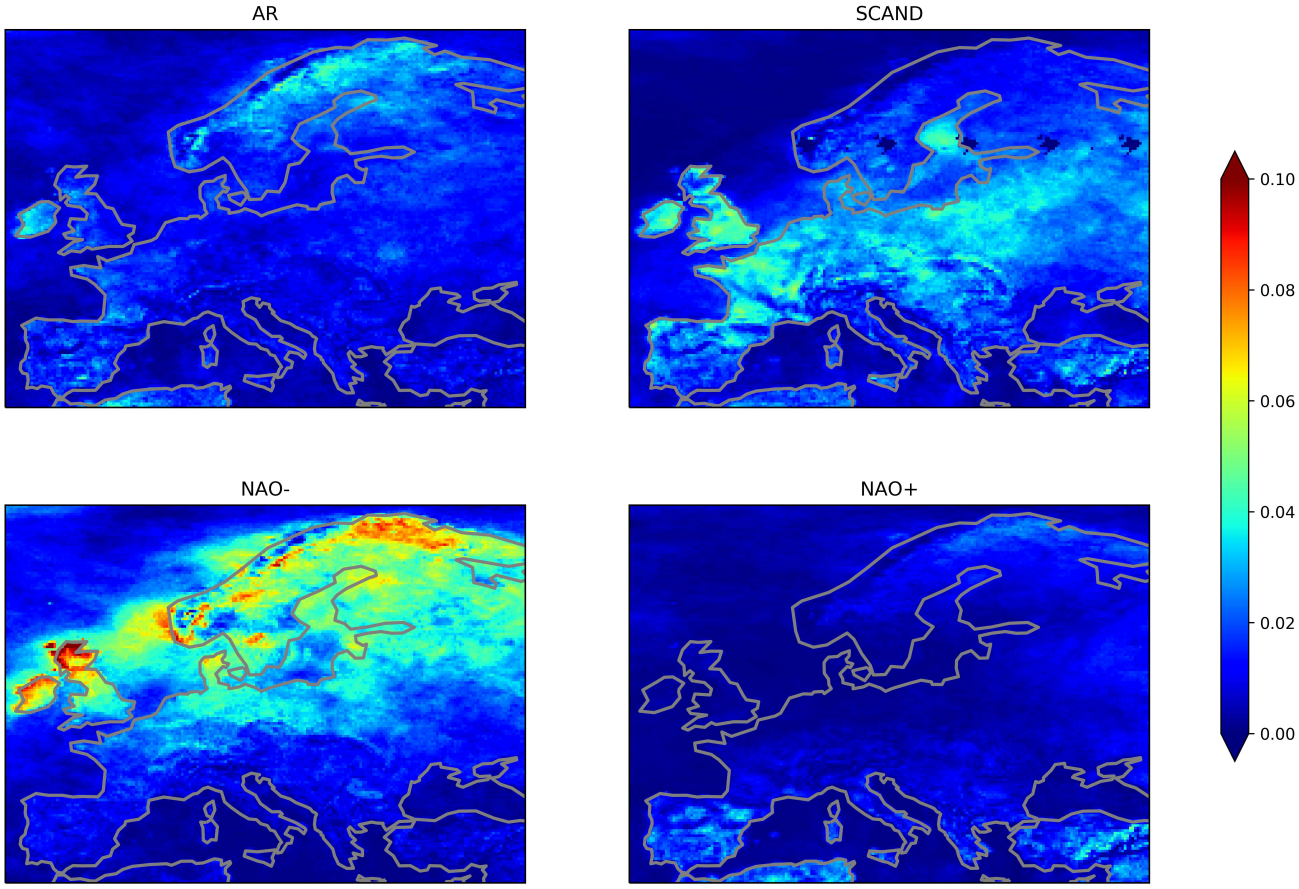

Figure C.21: Joint probabilities from copula model for each . Colours show the probabilities of occurrence of low temperature and low wind speed events computed with the copula model for minimum temperatures and daily maximum wind speeds below the 10th percentile. Area of study: 35N-72N, 11W-40E. Figure based on ERA5 data (NDJFM, 1979–2017).