

LETTER • OPEN ACCESS

A deep learning based classification of atmospheric circulation types over Europe: projection of future changes in a CMIP6 large ensemble

To cite this article: M Mittermeier *et al* 2022 *Environ. Res. Lett.* **17** 084021

View the [article online](#) for updates and enhancements.

You may also like

- [SMILES-X: autonomous molecular compounds characterization for small datasets without descriptors](#)
Guillaume Lambard and Ekaterina Gracheva
- [First Observation of the MeV Gamma-Ray Universe with Bijective Imaging Spectroscopy Using the Electron-tracking Compton Telescope on Board SMILE-2+](#)
Atsushi Takada, Taito Takemura, Kei Yoshikawa et al.
- [Changes in precipitation variability across time scales in multiple global climate model large ensembles](#)
Raul R Wood, Flavio Lehner, Angeline G Pendergrass et al.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
13 April 2022REVISED
21 June 2022ACCEPTED FOR PUBLICATION
12 July 2022PUBLISHED
27 July 2022

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

A deep learning based classification of atmospheric circulation
types over Europe: projection of future changes in a CMIP6 large
ensembleM Mittermeier^{1,*}, M Weigert², D Rügamer², H Küchenhoff² and R Ludwig¹¹ Department of Geography, Ludwig-Maximilians-Universität München (LMU), Munich, Germany² Statistical Consulting StaBLab, Department of Statistics, Ludwig-Maximilians-Universität München (LMU), Munich, Germany

* Author to whom any correspondence should be addressed.

E-mail: m.mittermeier@lmu.de**Keywords:** circulation type classification, internal climate variability, deep learning, SMILE, Großwetterlagen, SMHI-LENS

Abstract

High- and low pressure systems of the large-scale atmospheric circulation in the mid-latitudes drive European weather and climate. Potential future changes in the occurrence of circulation types are highly relevant for society. Classifying the highly dynamic atmospheric circulation into discrete classes of circulation types helps to categorize the linkages between atmospheric forcing and surface conditions (e.g. extreme events). Previous studies have revealed a high internal variability of projected changes of circulation types. Dealing with this high internal variability requires the employment of a single-model initial-condition large ensemble (SMILE) and an automated classification method, which can be applied to large climate data sets. One of the most established classifications in Europe are the 29 subjective circulation types called *Grosswetterlagen* by Hess & Brezowsky (HB circulation types). We developed, in the first analysis of its kind, an automated version of this subjective classification using deep learning. Our classifier reaches an overall accuracy of 41.1% on the test sets of nested cross-validation. It outperforms the state-of-the-art automatization of the HB circulation types in 20 of the 29 classes. We apply the deep learning classifier to the SMHI-LENS, a SMILE of the Coupled Model Intercomparison Project phase 6, composed of 50 members of the EC-Earth3 model under the SSP37.0 scenario. For the analysis of future frequency changes of the 29 circulation types, we use the signal-to-noise ratio to discriminate the climate change signal from the noise of internal variability. Using a 5%-significance level, we find significant frequency changes in 69% of the circulation types when comparing the future (2071–2100) to a reference period (1991–2020).

1. Introduction

Large-scale atmospheric circulation in the mid-latitudes drives European weather and climate through the westerly jet stream and high- and low pressure systems originating from it (Woollings *et al* 2010, Huguenin *et al* 2020). Classifying the highly dynamic atmospheric circulation into discrete classes has been a key effort in synoptic climatology to gain a better understanding of the linkages between atmospheric forcing and surface conditions. Various different circulation type classifications exist. These can be categorized as subjective (manual), hybrid (mixed) or objective (automated/computer-assisted). Every

classification consists of two steps: the class definition and the allocation of pressure fields to these classes. For subjective classifications, the classes are manually defined by experts *a priori* to the assignment step, which is then also carried out manually. Hybrid methods are based on subjective class definitions with automatized assignment steps. For objective methods, in contrast, the entire procedure is carried out in a numerical, automated way (Huth *et al* 2008).

One of the most established classification schemes in Europe comprises 29 circulation types called *Grosswetterlagen* by Hess & Brezowsky (HB circulation types; Hess and Brezowsky 1952). Werner and Gerstengarbe (2010) published a revised catalog that

covers the period from 1881–2009 and provides daily information on the HB circulation types. The catalog is constantly updated by the German Weather Service (DWD). Even though the subjectiveness of the HB circulation types involves considerable disadvantages in terms of inconsistencies and ambiguous class assignments, the main advantages of this classification are its intuitive naming convention and its high quality, e.g. for the description of climate elements especially over Central Europe (Sýkorová and Huth 2020). The main benefits compared to an automated classification (e.g. by cluster analysis) are the abilities to describe real synoptic features and to also capture rare but relevant synoptic types (e.g. a specific type of blocking anticyclones; James 2006b).

Due to these reasons the HB circulation types have been widely used for applications that study the connection between atmospheric circulation and extreme events (Sýkorová and Huth 2020); this includes heavy rainfall (Minářová *et al* 2017), floods (Petrow *et al* 2009), extreme temperatures (Sulikowska and Wypych 2020) and heat waves (Hoy *et al* 2020). The impact of the HB circulation types on weather exposed sectors like renewable energies has also been investigated (Drücke *et al* 2021). Sulikowska and Wypych (2020) discovered that most of the hot days of the exceptionally hot summer in Europe in 2019 occurred in connection with only four dominant HB circulation types. Petrow *et al* (2009) identified a few circulation types that trigger the majority of flood events in Germany and found that some of these types significantly increased during the period from 1952 to 2002. By analyzing historic trends, (Hoffmann and Spekat 2021) found that wet- and dry HB circulation types have significantly changed in frequency and duration from 1961 to 2018, and suggest that changes in European rainfall patterns are largely caused by dynamical changes of circulation types.

Because of the connections between extreme events or climate variables of interest and driving circulation types, it is highly relevant to understand future changes in the occurrence of circulation types in the context of climate change. Huguenin *et al* (2020) studied dynamic changes of large-scale atmospheric circulation types that are based on the HB circulation types and summarized them in ten groups of atmospheric flow (Beck *et al* 2007). Using a multi-model ensemble, they found no clear future trend in frequency or persistence of the circulation types, and explained this with the large influence of both internal variability and model spread between different climate models (Huguenin *et al* 2020). Due to its dynamic nature, the large-scale atmospheric circulation is highly variable. For the detection of future changes in circulation patterns it is therefore essential to consider the range of internal variability of the climate system (Vautard *et al* 2016). While HB circulation types have been widely used in conjunction with historic data, only James (2006a) and Ringer

et al (2006) have examined future changes of all 29 HB circulation types in climate models. They used an automated (hybrid) version of the HB circulation types developed by James (2006b). This automated version uses climate mean composite plots (separately for winter and summer) of all 29 circulation types based on daily mean fields of sea level pressure (slp) and geopotential height at 500 hPa (z500). A specific day in the climate model is assigned to the HB circulation type whose composite field has the highest correlation coefficient to the smoothed mean pressure field of the given day. Using this method, (James 2006a) found no clear trends for future circulation changes in HadGEM1 climate model runs and attributed this to the high interannual variability. James (2006a) states that a large database is needed in order to derive robust statements about changes in the European circulation patterns.

In summary, HB circulation types are widely used to study extreme events and weather exposed sectors in Europe, but there is a lack of knowledge regarding future changes of these circulation types due to internal variability.

In this paper, we introduce a new automated (hybrid) version of the classification of *Grosswetterlagen* by Hess and Brezowsky (1952) using deep learning. The code of this classification method is published open-source (Mittermeier *et al* 2022; see data availability statement) and enables the classification of HB circulation types in large climate ensembles. The application to a single-model initial-condition large ensemble (SMILE) allows us to investigate changes in the occurrence of the 29 HB circulation types under climate change conditions while considering the highly relevant influence of internal variability. A SMILE contains several simulations (members) of one climate model that only differ in their initialization. Thus, the members are equally likely realizations of the future climate and span the uncertainty range of internal variability introduced by small differences in the initial conditions (Deser *et al* 2012, Maher *et al* 2021). Deep learning is the state of the art method for visual pattern recognition, which has been applied to different climate pattern classification and detection problems (Liu *et al* 2016, Racah *et al* 2016, Kurth *et al* 2017, Huntingford *et al* 2019, Mittermeier *et al* 2019). Deep neural networks are capable of learning complex non-linear relationships in the data and are considered to have a high potential for solving challenging tasks in atmospheric sciences that involve vast amounts of spatio-temporal data (Liu *et al* 2016, Rolnick *et al* 2019). We train a deep learning classifier to distinguish the 29 circulation types based on the classification decisions in the long historic record of subjective classifications carried out by experts. It then provides an automated version of the HB circulation type classification that comes with low computational costs and is appropriate for handling large data sets like SMILES.

2. Data and methods

2.1. Training data set

We train our deep learning classifier on historic examples of HB circulation types for the period from 1900 to 1980. The supervised training process is based on two data components. First, the catalog of *Grosswetterlagen* over Europe by Hess & Brezowsky (Werner and Gerstengarbe 2010) contains a list of daily class affiliations for the 29 HB circulation types since 1900 derived from a manual classification of observed atmospheric pressure constellations. We use the catalog's class affiliations as labels for the training of our deep neural network. Table 1 lists the 29 circulation patterns with their acronyms and full names. The second data component is the ERA-20C reanalysis by the European Centre for Medium-Range Weather Forecasts (Poli *et al* 2016) covering the period from 1900 to 2010. This data contains the spatial atmospheric pressure patterns that match the labels of the catalog and are interpreted as images according to their pixel-wise structure. We use the variables slp and z500 in a 5° spatial resolution over a domain covering Europe and parts of the North Atlantic (30° N–75° N, –65° O–45° O) based on Werner and Gerstengarbe (2010). Due to an implausible sudden discontinuity of the labels of the catalog that starts around the mid-1980s with an artificial increase in circulation type persistence (Kučerová *et al* 2017), the period from the year 1980 onward is excluded and only the consistent data from 1900 to 1980 is used for training. The training database contains 29 585 training examples of daily, historic HB circulation types. Figure 1 illustrates the typical air pressure constellations for each of the 29 classes for slp and z500.

2.2. Network architecture and configuration

Our classification approach builds upon the image-like structure of the circulation patterns and uses a convolutional neural network. Its architecture is an adaptation of the model provided by Liu *et al* (2016) in the context of weather pattern detection and consists of two convolutional layers, a dropout layer and two fully connected layers. In the convolutions, we use two individual channels for the climate parameters (slp and z500). Based on the original definition by Hess and Brezowsky (1952), the circulation types have to last for at least three days. This is why we apply transition smoothing as a post-processing step and smooth out class predictions that last for less than three days (details in the appendix; section ‘Transition smoothing’). The model is trained using Adam optimization (Kingma and Ba 2014) with a batch size of 128, for 35 epochs and early stopping with a patience of six epochs. Hyperparameter tuning for learning rate and dropout rate is performed using

Bayesian optimization (Snoek *et al* 2012). The performance of the model is evaluated using the overall accuracy and the macro F1-score (see appendix: equations (A.1)–(A.3)), which takes the average of the class-specific F1-scores and has a value range from 0 to 1 (Opitz and Burst 2021). To obtain reliable and robust performance estimates, we apply nested cross-validation (Cawley and Talbot 2010) with an inner for-loop for model tuning and an outer for-loop to split off independent test sets. We use folds that contain ten years each, i.e. eight outer folds (test sets) and seven inner folds (model tuning). For each inner fold and its best hyperparameter set, we train five networks to account for different random weight initializations. The performance metrics (e.g. overall accuracy, F-scores) quantifying the performance of the deep learning classifier are derived by taking the average over the eight outer test sets and five networks. This results in robust performance metrics derived from balanced and independent test sets. To account for the time series nature of circulation patterns, training examples from the same year are required to be in the same cross-validation fold.

To derive the final weights of a trained deep neural network that can be used for applications on new data (e.g. the SMHI-LENS), all available training examples from 1900 to 1980 are used for training without splitting off test sets (Hastie *et al* 2009). Model tuning (the inner loop of the nested cross-validation) is applied again to find the best hyperparameter configuration before training with all data.

2.3. Uncertainty assessment

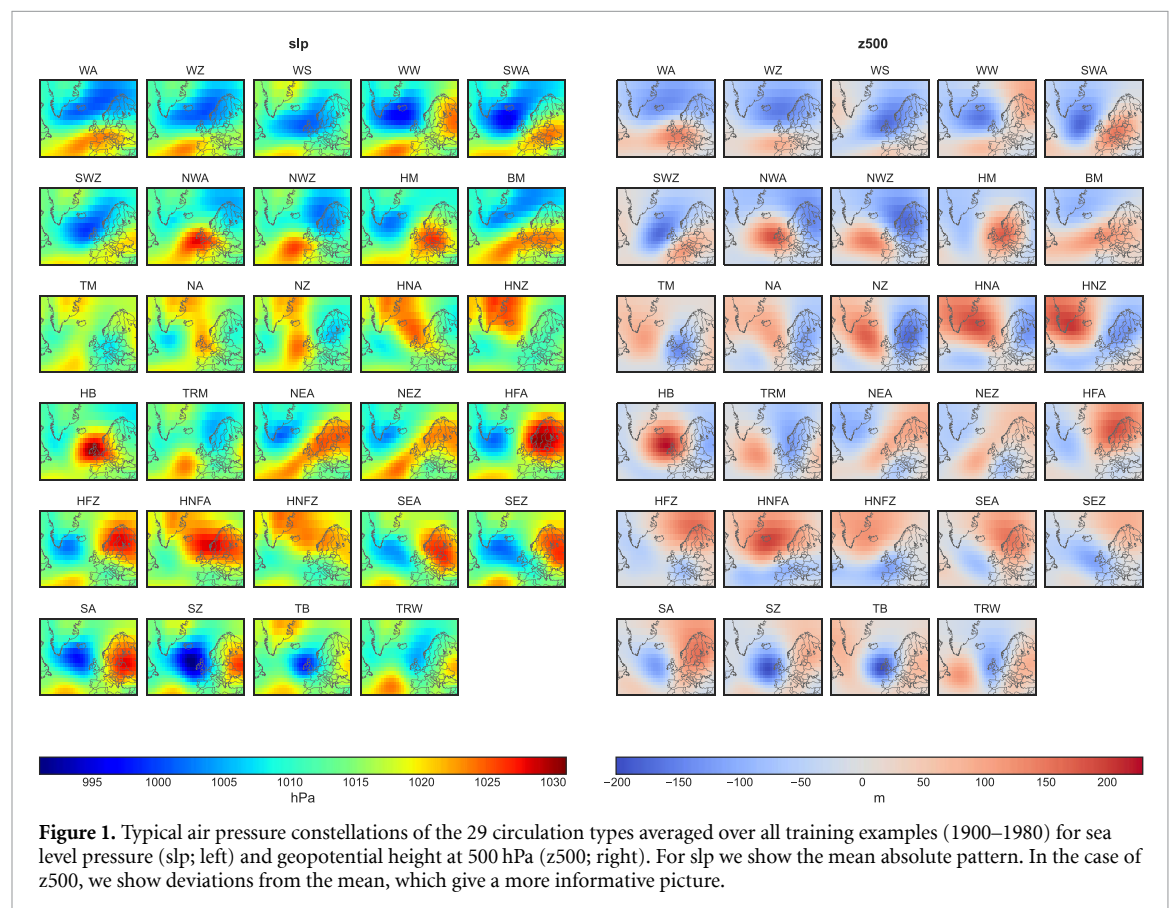
Due to their complexity, deep neural network training and their predictions are subject to uncertainty. In order to quantify the uncertainty of our deep learning classifier, we use a deep ensemble (Lakshminarayanan *et al* 2017) by generating 30 networks based on different random weight initializations while all other settings (e.g. hyperparameter configurations) are kept stable. Using this approach, we can quantify the variance of predictions and generate more robust class affiliations by applying all 30 networks to the data and calculating a weighted average prediction (Krogh and Vedelsby 1994). The weighted average considers the trust in each of the 30 networks as quantified by the F1-score. Instead of applying only a single final model, we apply the deep ensemble of 30 networks to new data.

2.4. Climate ensemble: SMHI-LENS

The deep ensemble introduced in section 2.3 is furthermore applied to the climate ensemble SMHI-LENS (Wyser *et al* 2021). SMHI-LENS is a SMILE of the Swedish Meteorological and Hydrological Institute (SMHI), with the EC-Earth model (version 3.3.1;

Table 1. List of the 29 circulation patterns with their acronyms, original German name and translated English name based on James (2006b).

Acronym	Original name (German)	Translated name (English)
WA	Westlage, antizyklonal	Anticyclonic Westerly
WZ	Westlage, zyklonal	Cyclonic Westerly
WS	Südliche Westlage	South-Shifted Westerly
WW	Winkelförmige Westlage	Maritime Westerly (Block E. Europe)
SWA	Südwestlage, antizyklonal	Anticyclonic North-Westerly
SWZ	Südwestlage, zyklonal	Cyclonic South-Westerly
NWA	Nordwestlage, antizyklonal	Anticyclonic North-Westerly
NWZ	Nordwestlage, zyklonal	Cyclonic North-Westerly
HM	Hoch Mitteleuropa	High over Central Europe
BM	Hochdruckbrücke (Rücken) Mitteleuropa	Zonal Ridge across Central Europe
TM	Tief Mitteleuropa	Low (Cut-Off) over Central Europe
NA	Nordlage, antizyklonal	Anticyclonic Northerly
NZ	Nordlage, zyklonal	Cyclonic Northerly
HNA	Hoch Nordmeer-Inland, antizyklonal	Icelandic High, Ridge C. Europe
HNZ	Hoch Nordmeer-Inland, zyklonal	Icelandic High, Trough C. Europe
HB	Hoch Britische Inseln	High over the British Isles
TRM	Trog Mitteleuropa	Trough over Central Europe
NEA	Nordostlage, antizyklonal	Anticyclonic North-Easterly
NEZ	Nordostlage, zyklonal	Cyclonic North-Easterly
HFA	Hoch Fennoskandien, antizyklonal	Scandinavian High, Ridge C. Europe
HFZ	Hoch Fennoskandien, zyklonal	Scandinavian High, Trough C. Europe
HNFA	Hoch Nordmeer-Fennoskandien, antizykl.	High Scandinavia-Iceland, Ridge C. Europe
HNFZ	Hoch Nordmeer-Fennoskandien, zyklonal	High Scandinavia-Iceland, Trough C. Europe
SEA	Südostlage, antizyklonal	Anticyclonic South-Easterly
SEZ	Südostlage, zyklonal	Cyclonic Southerly
SA	Südlage, antizyklonal	Anticyclonic Southerly
SZ	Südlage, zyklonal	Cyclonic Southerly
TB	Tief Britische Inseln	Low over the British Isles
TRW	Trog Westeuropa	Trough over Western Europe



Döscher *et al* 2022) and 50 members. The SMHI-LENS follows the protocol of the Coupled Model Intercomparison Project phase 6 (CMIP6). We chose the SMHI-LENS for its high number of members and the high performance of the EC-Earth3 model in reproducing daily sea-level pressure circulations types. Cannon (2020) compared 15 general circulation models with two reanalysis data sets. The EC-Earth3 was found to be one of the best performing CMIP6 models in terms of reproducing frequency and persistence of circulation types under the consideration of internal variability, especially over Europe (Cannon 2020). The SMHI-LENS is available for the period from 1970 to 2100 for four different scenarios with a 0.7° spatial resolution. It uses the macro initialization method for the generation of its ensemble members. We use the high-emission climate scenario SSP37.0 and a daily resolution. The data is clipped to the Europe-North-Atlantic domain (see section 2.1) and regridded to the 5° grid used during training of the deep learning classifier by means of bilinear interpolation. Frequencies of occurrence of circulation patterns are compared for two 30 year periods, a far future horizon from 2071 to 2100, and the reference period from 1991 to 2020. The signal-to-noise ratio (S/N-ratio) and its significance is calculated according to Aalbers *et al* (2018) using a two-sided *t*-test. The S/N-ratio states, if the forced response (ensemble averaged frequency change) exceeds the noise (standard deviation of the ensemble). As we simultaneously conduct hypothesis tests for all 29 circulation types, adjustments for multiple testing are needed to reduce the risk of incorrectly rejecting null hypotheses. We apply the method of Benjamini and Hochberg (1995) to control the false discovery rate, i.e. the proportion of incorrectly significant findings among all significant findings, for the chosen alpha level of 0.05.

3. Results

3.1. Method evaluation

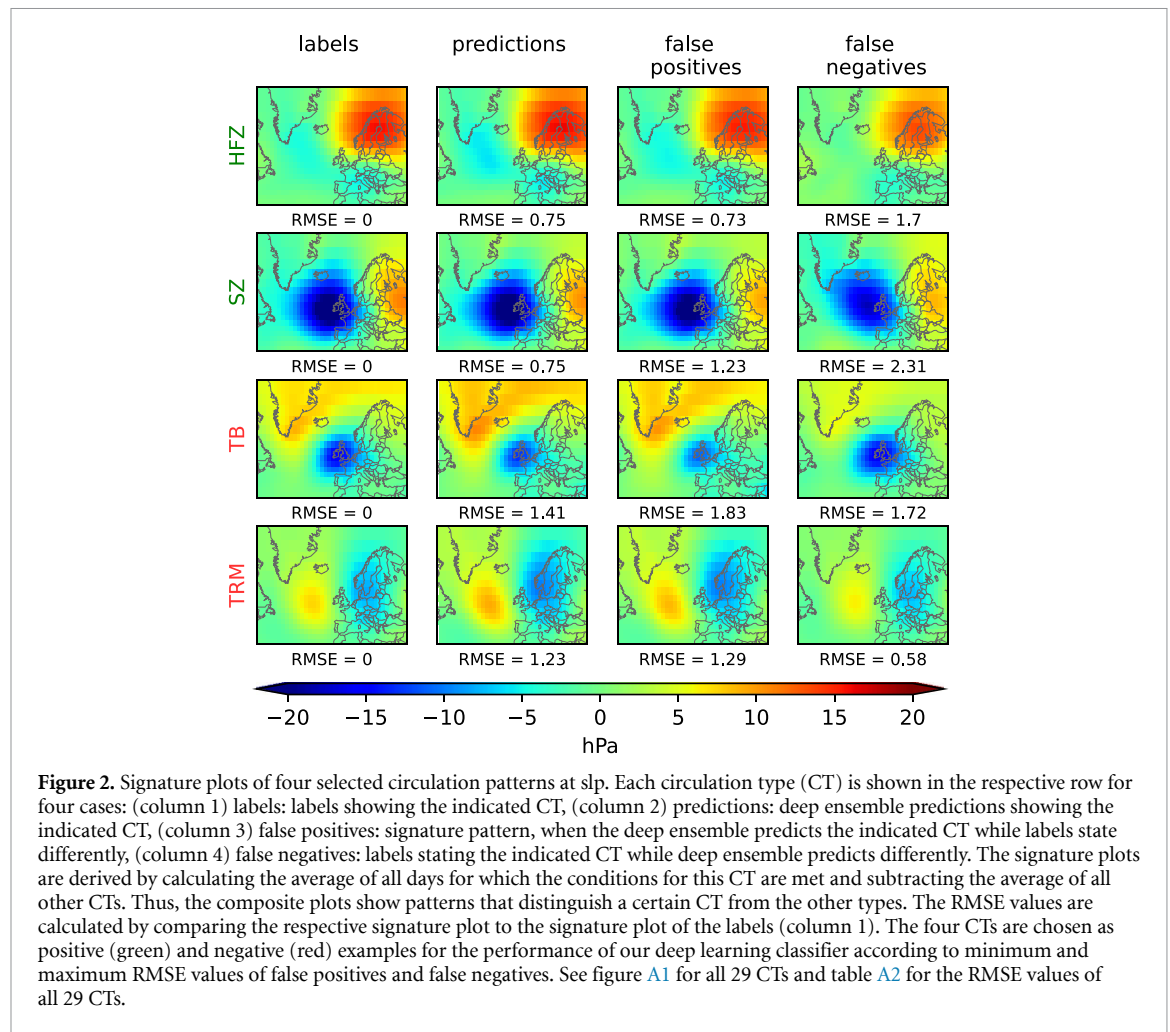
To evaluate the performance of our method, the daily class affiliations of the original HB circulation type catalog (Werner and Gerstengarbe 2010) are compared to the class predictions of the deep learning classifier. On the outer folds of the nested cross-validation, we obtain a macro F1-score of 39.3 and an overall accuracy of 41.1%. The class-specific F1-scores are given in the second column of table 2.

Table 2 further shows the performance measures of the method by James (2006b). While the class-specific F1-scores for our deep learning classifier are derived from independent test sets during nested cross-validation based on ERA-20C reanalysis data for the period 1900–1980, the class-specific F1-scores for the method by James (2006b) are based on ERA-40 reanalysis data (Uppala *et al* 2005) for the time period from September 1957 to August 2002. While a direct comparison between the two approaches is

Table 2. Comparison of class-specific F1-scores of our Deep Learning classifier (DL) evaluated during nested cross-validation (CV) on ERA-20C and comparison to the classification method of James (2006b) on ERA-40. The best results are highlighted in bold in order to facilitate the comparison of the two methods. The overall accuracy and macro F1-Scores are given in the last two rows.

Circulation type	F1-score DL CV	F1-score James
WA	44.6	40.04
WZ	47.08	52.5
WS	45.39	34.89
WW	37.7	29.91
SWA	35.36	36.44
SWZ	30.86	39.44
NWA	38.88	33.51
NWZ	37.07	43.28
HM	51.24	43.07
BM	47.29	37.88
TM	37.23	36.96
NA	24.85	15.82
NZ	44.32	41.31
HNA	45.57	45.55
HNZ	27.11	36.53
HB	50.99	44.78
TRM	27.86	39.35
NEA	41.44	29.74
NEZ	33.12	27.12
HFA	45.32	40.94
HFZ	24.81	32.85
HNFA	33.35	43.21
HNFZ	34.02	33.06
SEA	38.09	27.25
SEZ	37.93	31.01
SA	39.84	33.89
SZ	38.19	26.57
TB	42.11	37.7
TRW	29.34	37.64
Macro F1-score	38.3	36.28
Overall accuracy	41.1	39.1

thus not exact, contrasting both approaches is still valid under the assumption that both data sets are representative for the underlying distribution and class ratios. Given the length of both observation periods, this seems to be a reasonable assumption. In addition, our nested cross-validation approach can be considered robust without the risk of drawing an overly optimistic comparison in favor of our method. As table 2 shows, the deep learning classifier outperforms the method by James (2006b) in 20 of the 29 classes. The overall accuracy of our deep learning method is 41.1% (macro F1-score: 38.3) and 39.1% (macro F1-score: 36.28) for James (2006b). For the circulation patterns WS, NEA, SEA and SZ, the performance of the deep learning classifier is more than 10% higher, while the approach by James (2006b) works especially well for TRM and HNFA. The confusion matrix showing the average classifications of our deep learning classifier on the test sets during cross-validation is given in table A1 in the appendix. Most of the misclassifications



occur between pairs of anticyclonic- and cyclonic circulation.

Our deep learning classifications are compared to the HB circulation type catalog in respect to the frequency distribution of the classes (see [appendix](#), figure A2(a)). The network reproduces the relative order of the classes well, but clearly underestimates the class WZ. Classes HM and BM are also underestimated by the network, while it overestimates class WS. The climate ensemble SMHI-LENS reproduces the circulation types well (see [appendix](#), figure A2(b)). Except for BM, for which the climate model overestimates the frequency, all boxplots cover or intersect with the frequencies in ERA-20C reanalysis data.

Figure 2 evaluates the ‘synoptic performance’ (Verdecchia *et al* 1996) of our deep learning classifier for four selected cases. Figure A1 in the [appendix](#) shows all 29 circulation types. The signature plots in figures 2 and A1 are derived by taking the average field of a certain class and subtracting the average field of all other classes from it. This shows which synoptic characteristics distinguish a single class from the other classes. Signature plots are given for four different cases: labels (column 1), predictions (column 2), false positives (column 3) and false negatives

(column 4). If columns 1 and 2 are very similar, the average signature of the deep learning prediction agrees well with the average signature of the labels as derived from the original HB catalog. The four selected cases in figure 2 show two positive (green) and two negative (red) examples for the performance of our deep learning classifier. Signature plots allow a visual comparison of the spatial patterns of the circulation types. This reveals for example if the intensity of low pressure system is on average weaker in the network predictions compared to the labels (as it is the case for TB) or if the spatial extent of a synoptic feature is overestimated (like for TRM). A further noteworthy insight is given if there is good agreement between column 1 and 3 (as it is the case for HFZ), which can indicate a misclassification in the catalog. In this case, our deep learning method might correctly classify the situation, while the catalog labels disagree. Additionally, the difference between the signature plots is quantified using the root-mean-square error (RMSE) as metric (see equation (A.4) and table A2 in the [appendix](#); Müller *et al* 2020). In case of a perfect match between two spatial patterns, the RMSE has a value of zero. The four examples in figure 2 are chosen according to the minimum and maximum RMSE values for false positives and false negatives.

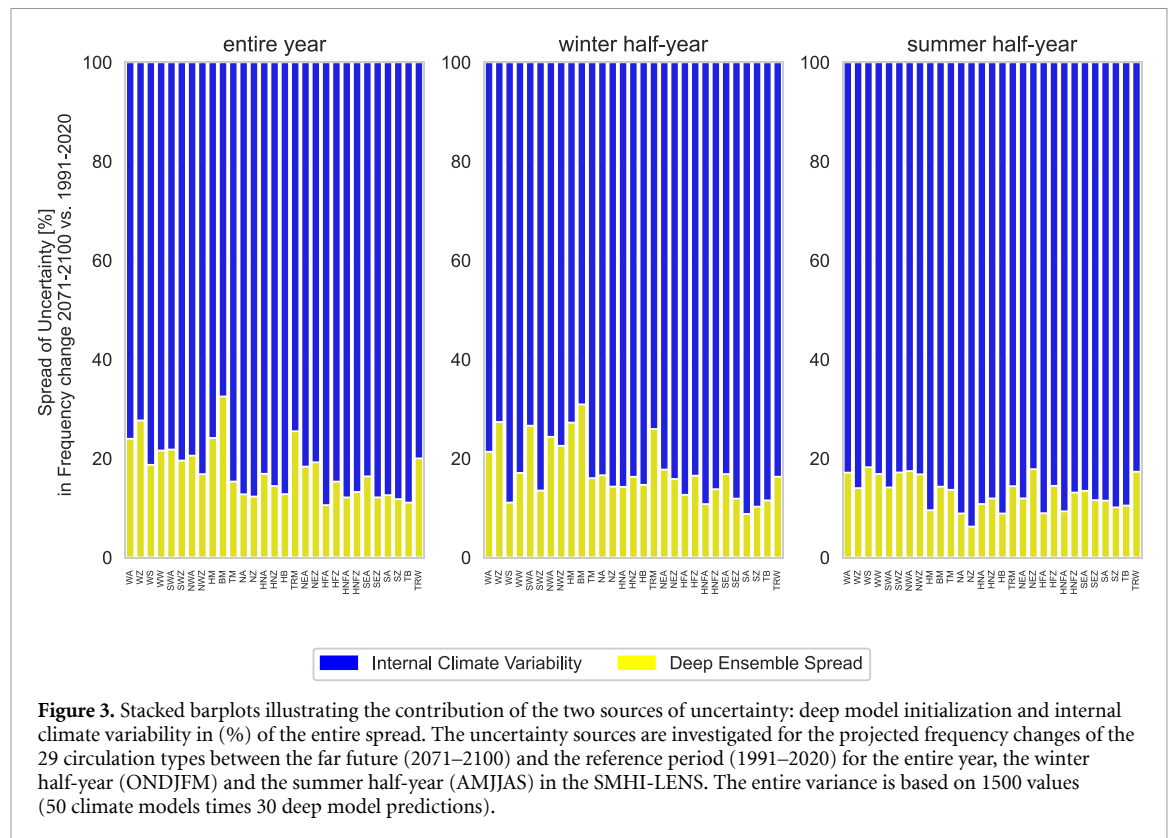


Figure 3. Stacked barplots illustrating the contribution of the two sources of uncertainty: deep model initialization and internal climate variability in (%) of the entire spread. The uncertainty sources are investigated for the projected frequency changes of the 29 circulation types between the far future (2071–2100) and the reference period (1991–2020) for the entire year, the winter half-year (ONDJFM) and the summer half-year (AMJJAS) in the SMHI-LENS. The entire variance is based on 1500 values (50 climate models times 30 deep model predictions).

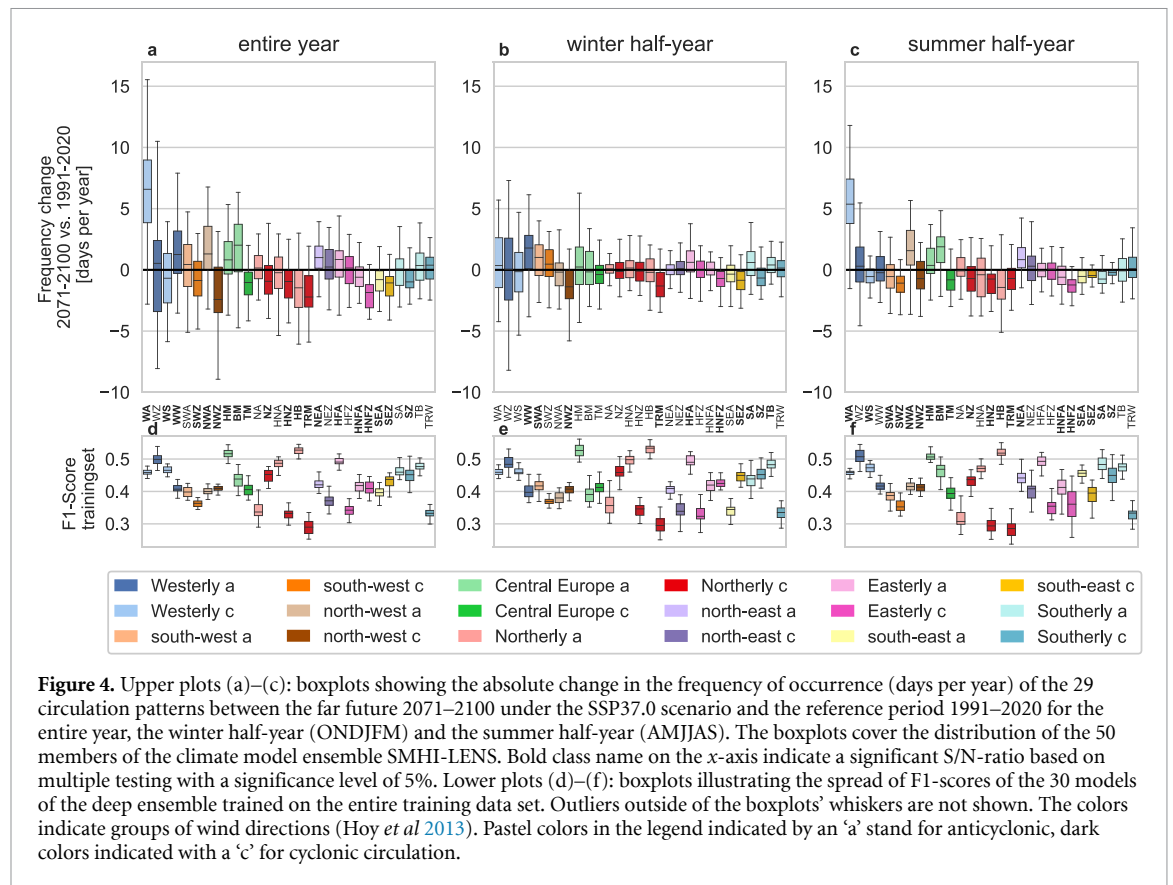
The RMSE when comparing the signature plot of the predictions (column 2) with the signature plot of the labels (column 1) has on average over all 29 circulation types a value of 0.89 (see table A2 in the appendix). For the false positives, the RMSE is slightly higher with on average 1.09. This matches to the visual comparison of the signature plots: apart from slight differences for some classes, the spatial patterns in column 3 agree generally well with column 1 (see figure A1). Exceptions for which the false positives differ the most from the signature plot of the labels are: TB, for which the extent of the low pressure system is too small (RMSE = 1.83), SA, whose low pressure system is too weak (RMSE = 1.69) and HNA and HNZ, for which in both cases the low pressure system in the South is too strong (RMSE = 1.41 for both). In other cases with low RMSE values and good agreement of the spatial patterns, the signature plots are an indication to question the choice in the subjective catalog. For example for HFZ (RMSE = 0.73), NEA (RMSE = 0.76) and WA (RMSE = 0.78). Although the deep learning classifier only considers the information provided in the data and subjective reasons for the label classification are not available, a certain level of arbitrariness in the catalog has already been recognized before by James (2006b). Column 4 reveals the false negatives of the deep learning classifier. Here, the average RMSE value over all 29 circulation types is 1.28. In some cases (e.g. SZ, WS, HNFA, TB and HFZ), the false negatives clearly differ from the signature patterns of the labels in column 1. In these cases, the catalog labels may be questioned and predictions

of our approach seem plausible—at least based on the objective slp and z500 data.

Figure 3 depicts the uncertainty obtained by our deep ensemble (30 members) compared to the internal climate variability (50 members) by plotting stacked barplots of the percentages of the total uncertainty (50 climate model members times 30 deep ensemble members) with the respective attribution to these two sources. The network uncertainty range lies at 11%–33% for the entire year. It is larger in winter and smaller in summer. Note that for the deep learning part this does not take the variability of hyperparameter tuning into account. With regard to typical climate modeling uncertainties (Hawkins and Sutton 2011), the influence of climate model choice and scenario choice is not considered.

3.2. Future changes

We apply the weighted deep ensemble to the SMHI-LENS with its 50 members to quantify the spread of internal variability for future frequency changes of the 29 circulation patterns between 2071–2100 and the reference period (1991–2020). Figure 4 shows absolute frequency changes and the spread of internal variability for all circulation types for the entire year as well as the winter and summer half-year illustrated by boxplots. Relative frequency changes are illustrated in figure A3. Significant changes in terms of the S/N-ratio are indicated with bold class names. Table A3 shows the complete list of S/N-ratio values for the absolute frequency changes. For most circulation types, the boxplots intersect with the horizontal line



at zero and the members disagree in the sign of the trend. Overall, absolute changes are small and lie within a range of ± 5 days for most circulation types. This finding is in line with Huguenin *et al* (2020), who find small changes of ± 4 days per season in a multi-model ensemble for ten groups of circulation patterns that are based on the 29 HB circulation types. Note that for the circulation types TM-TRW (in the presented order), which have small absolute frequencies, changes of ± 5 days can still mean high relative changes of around $\pm 50\%$ (see appendix, figure A3). For some circulation types, single members outside the interquartile range show relative changes of $> 50\%$, especially in the winter half-year. Different to the studies by James (2006a) and Ringer *et al* (2006), who analyzed all 29 circulation types in the climate model HadGEM1 and have not found significant frequency changes, our analysis of a SMILE allows to identify significant frequency changes due to climate change despite the high spread of internal variability.

In figure 4(d)–(f), we plot the class-specific F1-scores of our deep learning classifier and their range throughout the deep ensemble. This allows to take into account the quality of predictions for each class. Reliable statements can be made for the circulation patterns WA, WZ, WS, HM, HNA, HB, HFA, SA, and TB throughout the entire year. The clearest absolute climate trend is found for the anticyclonic westerly circulation (WA), which shows an increasing trend

for the entire year (and the summer half-year) with a median of 6.6 days per year (summer: 5.4 days per year) and a S/N-ratio of 1.5 (summer: 1.7). For WA, the climate change signal clearly exceeds the noise of internal variability. The increasing winter trends of HFA and TB are also significant, as well as decreasing summer trends of WS, HB and SA and the increasing summer trend for HM.

In general, we find a decreasing trend for south-easterly circulations (SEA and SEZ) in both summer and winter (trends are significant except for SEA in winter), although their reliability based on F1-scores fluctuates seasonally. For winter, this goes along with the findings by Herrera-Lormendez *et al* (2021), who have detected a decreasing trend for south-easterly circulations from the Jenkinson–Collison classification using four members of EC-Earth3 under SSP58.5. The classification by Jenkinson and Collison (1977) is an automated version of the subjective Lamb catalog developed for the British Isles. Herrera-Lormendez *et al* (2021) applied this classification to Europe and distinguished 11 circulation types. Our results also support the findings of Herrera-Lormendez *et al* (2021) for the increasing summer trend of north-easterly circulations (in our case significant for NEA) and the decreasing summer trend for Northerlies (in our case significant for NZ and HB).

Our results make clear that the spread of internal variability is tremendous and it is difficult to derive

systematic changes of circulation patterns grouped by their wind directions. Despite the high internal variability, the results of the S/N-ratio are very clear, showing a significant change in 69% of the classes for the total year, 34% for the winter and 69% for the summer half-year.

4. Conclusion

In this study, we introduced a new automated classification method for the 29 circulation types defined by Hess and Brezowsky (1952) using deep learning. Our method shows the potential of deep learning in circulation type classification and outperforms the state-of-the-art method of James (2006b) in 20 of the 29 classes. We applied the deep learning classifier to a SMILE of the CMIP6 generation, the SMHI-LENS, which comprises 50 members of the EC-Earth3 general circulation model. Our study is the first one that analyzes future frequency changes of all 29 circulation patterns in a SMILE. In contrast to previous studies on climate change impacts on the HB circulation types (Ringer *et al* 2006, James 2006b), we can thus identify significant frequency changes despite the high range of internal variability.

A better understanding of climate change impacts on the European circulation patterns is of high societal relevance because of their direct influence on our daily weather and the strong relation to extreme events like heavy rainfall (Minářová *et al* 2017), floods (Petrov *et al* 2009), hot days (Sulikowska and Wypych 2020) and heat waves (Hoy *et al* 2020). Our results show an immense spread of internal variability when investigating future frequency changes of the circulation patterns in the SMHI-LENS under the SSP37.0 scenario. Despite the high spread of internal variability, our results of the S/N-ratio show significant ($\alpha = 5\%$) absolute frequency changes for a high number of classes (69% of the classes for the entire year, 34% for the winter half-year and 69% for the summer half-year). This underlines the great benefit in using a SMILE when analyzing climate change effects on the highly dynamic large-scale atmospheric circulation over Europe. In absolute numbers, frequency changes lie in a range of ± 5 days for most circulation types, which agrees with the findings by Huguenin *et al* (2020). For the circulation types TM-TRW (in the presented order), which occur only on a few days per year, small absolute changes can still mean high relative changes (for some circulation types around $\pm 50\%$, for some members even $> 50\%$). The most distinct absolute change is found for Anticyclonic Westerlies (WA) with an increasing trend for the entire year with a median of 6.6 days per year and a S/N-ratio of 1.5. Here, the climate change signal clearly exceeds the noise of internal variability.

The classification results in section 3.1 show that our deep learning classifier can yield good predictions at low computational costs. This makes our method

advantageous for application to large climate data sets such as multi-model ensembles or SMILEs. Regarding the goal of reproducing the original subjective HB circulation types, it achieves higher performance measures than the method by James (2006b). For some classes, a larger part of the misclassifications of our deep learning classifier seem to be synoptically correct. The labels of the HB circulation type catalog (Werner and Gerstengarbe 2010) are subjective and hold inconsistencies and ambiguous class affiliations (James 2006b, Kučerová *et al* 2017). This means that the labels taken as ground truth hold a certain, unquantified human level error. Our findings suggest that this human level error might be substantial for some classes.

Our deep learning classifier is designed for the application to climate models, as this requires an automated version of the HB circulation type classification. It is not meant to replace a subjective continuation of the HB catalog and it is not suitable for this as long as the human level error is unquantified and there is potential to improve the performance of the classifier. A disadvantage of the deep learning approach is its potentially high variability, which can be caused by model uncertainty or too noisy data. Our evaluation shows that the variability of the deep learning method contributes up to 32.5% of the entire variance when applying our method to the SMHI-LENS. To deal with this uncertainty, we use a deep ensemble of 30 networks with different initializations and calculate a performance-weighted mean of this deep ensemble when applying the classifier on new data.

Besides quantifying the human level error in the labels, possible future research could evaluate further network architectures for an improvement of the deep learning performance. Considering the temporal development of circulation patterns by using a temporal-aware ConvLSTM architecture might improve the classification accuracy. Furthermore, a deep hidden Markov model could improve the performance by including the three-day-definition of HB circulation types directly in the training process. In order to evaluate the uncertainties in frequency changes coming from different climate models and forcing scenarios, a combination of multi-model as well as single-model ensembles under different forcing scenarios is desirable. The deep learning classifier introduced in this study can serve as valuable tool for the analysis of such a comprehensive data set.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/mmittermeier/Deep-learning-classification-of-atmospheric-circulation-types-over-Europe.git>. The link leads to a GitHub repository that contains our trained deep learning classifier and the

code to use it. The code is based on python (version 3.6). ERA-20C reanalysis data is derived from the European Centre for Medium-Range Weather Forecasts (ECMWF): www.ecmwf.int/. The SMHI-LENS is publicly available from the data portal of the Earth System Grid Federation (ESGF): <https://esgf-data.dkrz.de/>.

Acknowledgments

The work of M M is funded through the ClimEx project (www.climex-project.org) by the Bavarian State Ministry for the Environment and Consumer Protection, the work of M W and D R by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The provision of the digital Hess & Brezowsky catalog for the years 1900–2010 by the German Weather Service is highly appreciated.

Appendix

Transition smoothing

Our classifications must adhere to the definition that a circulation type lasts for at least three days. A transition-smoothing step ensures that this rule is respected by post-processing the time series of the network classifications. Firstly, circulation types that last for less than three days are identified. Next, these transitions are tested for *neighborhood consistency* and *transition membership*. Neighborhood consistency describes the situation if the same circulation type occurs before and after the transition. The transition days are then smoothed by assigning this type to it. In case of transition membership different circulation types occur before and after the transition. Here, the transition days obtain the class affiliation of the circulation type before or after, depending on which class has a higher predicted probability.

Table A1. Confusion matrix of our proposed smoothed approach, averaged over the test sets of nested cross-validation. Correctly classified classes are highlighted in bold.

Labels																																
	WA	WZ	WS	WW	SWA	SWZ	NWA	NWZ	HM	BM	TM	NA	NZ	HNA	HNZ	HB	TRM	NEA	NEZ	HFA	HFZ	HNEA	HNFZ	SEA	SEZ	SA	SZ	TB	TRW	Σ	Precision	
WA	102	62	1	3	4	2	10	8	22	26	0	1	1	0	0	1	3	1	1	0	0	0	0	0	0	0	0	0	1	3	253	0.40
WZ	13	195	12	5	2	6	1	11	4	2	0	1	2	1	0	0	6	1	0	0	0	0	0	0	0	0	0	2	5	272	0.72	
WS	1	51	67	3	0	8	0	4	1	1	4	0	0	1	3	0	6	0	0	0	0	0	1	0	3	0	1	6	5	170	0.40	
WW	4	24	3	42	2	4	1	3	6	6	1	0	0	0	0	0	3	2	3	2	1	0	0	2	2	3	2	2	6	125	0.33	
SWA	14	18	1	4	34	11	0	0	25	7	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	4	0	1	2	125	0.27	
SWZ	4	36	9	5	8	30	0	1	4	1	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	2	1	4	5	115	0.26	
NWA	14	8	0	1	0	0	58	18	12	20	0	1	4	2	1	12	3	2	2	0	0	0	0	0	0	0	0	0	1	159	0.37	
NWZ	10	56	2	2	0	0	15	67	2	4	2	1	7	1	0	1	21	1	1	0	1	0	0	0	1	0	0	1	3	198	0.34	
HM	8	4	1	1	4	1	5	1	142	23	0	1	0	2	0	4	0	2	1	7	1	1	0	2	0	3	0	0	1	216	0.66	
BM	15	5	0	3	1	0	7	2	25	104	0	0	1	1	0	2	3	3	3	3	0	1	0	1	0	2	0	0	2	187	0.56	
TM	0	8	4	1	0	0	0	3	0	1	34	1	6	0	4	0	10	1	3	1	1	1	5	0	1	0	1	4	6	95	0.36	
NA	2	4	0	0	0	0	3	1	4	1	0	12	7	12	3	3	1	2	2	1	0	1	1	0	0	0	0	0	0	59	0.20	
NZ	2	7	0	0	0	0	6	12	1	2	6	4	52	4	5	3	15	1	2	0	0	0	1	0	0	0	0	0	1	125	0.42	
HNA	1	4	1	0	0	0	2	0	11	3	1	6	3	54	5	10	1	1	1	2	0	3	2	2	0	0	0	1	0	116	0.47	
HNZ	0	5	3	0	0	1	1	0	1	1	7	2	7	10	17	1	3	0	1	0	0	1	7	1	0	0	0	1	2	72	0.23	
HB	1	1	0	0	0	0	19	4	13	11	0	2	2	8	0	65	2	5	2	2	0	1	0	0	0	0	0	0	0	141	0.46	
TRM	3	14	2	1	0	1	1	17	1	4	6	0	9	1	0	0	35	1	2	0	0	0	0	0	1	0	0	1	8	109	0.32	
NEA	1	2	0	2	0	0	4	1	10	10	1	1	1	3	0	5	1	43	13	10	2	2	1	1	0	1	0	0	1	114	0.38	
NEZ	1	2	1	1	0	0	5	1	1	5	4	0	2	1	1	2	5	10	26	2	2	1	2	0	1	0	0	0	1	79	0.33	
HFA	1	1	0	1	1	0	0	0	21	5	0	0	0	3	0	1	1	9	2	63	5	4	1	9	2	4	0	0	1	135	0.47	
HFZ	0	1	1	2	0	0	0	1	1	1	3	0	0	0	0	0	0	1	3	8	12	12	1	4	2	5	1	0	1	61	0.20	
HNEA	0	1	0	0	0	0	0	0	2	3	1	1	2	11	1	2	1	3	3	12	1	21	8	3	1	0	0	1	1	79	0.27	
HNFZ	0	3	1	0	0	0	0	0	1	1	7	0	2	4	7	0	1	0	2	4	2	6	24	3	2	0	0	1	2	74	0.32	
SEA	0	1	1	1	1	0	0	0	7	2	0	0	0	3	0	0	0	0	1	12	2	2	3	32	9	8	1	1	1	87	0.36	
SEZ	0	2	3	4	0	0	0	1	0	1	2	0	0	0	0	0	2	1	2	1	4	0	3	6	24	2	3	2	3	68	0.35	
SA	1	2	0	3	4	3	0	0	16	5	0	0	0	1	0	0	0	0	0	6	0	0	0	7	1	34	4	1	3	94	0.36	
SZ	0	1	5	4	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4	8	19	7	5	63	0.30	
TB	2	20	7	4	2	5	0	3	3	1	3	0	0	0	1	0	3	0	0	1	0	0	1	1	2	2	2	44	12	120	0.37	
TRW	4	18	3	3	1	3	1	4	2	4	2	0	1	0	0	0	12	0	1	0	1	0	0	1	1	1	1	7	32	104	0.31	
Σ	205	557	126	97	68	81	140	164	340	255	88	35	111	123	51	115	140	94	79	143	36	49	67	78	59	76	36	89	116	3616	—	
Recall	0.50	0.35	0.53	0.43	0.50	0.37	0.42	0.41	0.42	0.41	0.39	0.34	0.47	0.44	0.33	0.57	0.25	0.46	0.33	0.44	0.33	0.44	0.36	0.40	0.41	0.45	0.53	0.49	0.28	0.41	—	

Output

Table A2. Values of the root-mean-square error (RMSE) when comparing the composite plot of the predictions, the false positives and the false negatives with the composite plot of the labels for the 29 circulation types. RMSE values averaged over all 29 circulation types are printed in bold.

RMSE	WA	WZ	WS	WW	SWA	SWZ	NWA	NWZ	HM	BM	TM	NA	NZ	HNA	HNZ
Predictions	0.93	0.82	1.0	0.88	0.77	0.82	0.73	1.06	0.78	1.04	0.91	0.87	0.39	0.85	1.29
False positives	0.78	1.11	1.21	1.08	1.23	1.0	0.83	1.06	0.80	0.84	0.92	0.92	0.89	1.41	1.41
False negatives	1.25	0.84	2.10	1.42	1.62	1.24	0.83	1.36	0.87	1.12	1.03	1.50	0.80	1.04	1.3
RMSE	HB	TRM	NEA	NEZ	HFA	HFZ	HNFA	HNfZ	SEA	SEZ	SA	SZ	TB	TRW	∅
Predictions	0.66	1.23	0.58	0.89	0.89	0.75	0.74	0.66	0.99	0.91	1.11	0.75	1.41	1.14	0.89
False positives	1.11	1.29	0.76	0.96	1.41	0.73	0.84	0.94	0.98	1.22	1.69	1.23	1.83	1.27	1.09
False negatives	1.57	0.58	1.0	0.99	1.25	1.70	1.99	1.0	1.31	1.58	1.32	2.31	1.72	0.61	1.28

Equations for macro F1-score

The Macro F1-score (F_1) is calculated as the arithmetic mean of the class-specific F1-scores ($F1_x$) of all classes (see (A.3); Opitz and Burst 2021). n describes the number of classes (in our case: 29). The F1-Score is based on precision and recall. Precision (P ; see (A.1); adjusted based on Lewis *et al* 1996) is calculated based on items correctly assigned to the specific class (true positives (TP)), and non-class members put into the class (false positives (FP)). Recall (R ; see (A.2); adjusted based on Lewis *et al* 1996) on the other hand considers class members not assigned to this class by the deep learning classifier (false negatives (FN); Lewis *et al* 1996).

$$P = \frac{TP}{(TP + FP)} \quad (A.1)$$

$$R = \frac{TP}{(TP + FN)} \quad (A.2)$$

$$F_1 = \frac{1}{n} \sum_x F1_x = \frac{1}{n} \sum_x \frac{2P_x R_x}{P_x + R_x}. \quad (A.3)$$

Equation for RMSE

Equation for the calculation of the RMSE with I being the predicted image (in our case: signature plot of the deep learning classifier) and K being the reference image (in our case: signature plot of the labels). M are number of rows and N the number of columns of the pictures to compare. The RMSE thus compares the pixel-wise values of two images. A value of zero indicates a perfect match. Equation based on Müller *et al* (2020).

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=0, j=0}^{M-1, N-1} [I(i, j) - K(i, j)]^2}. \quad (A.4)$$

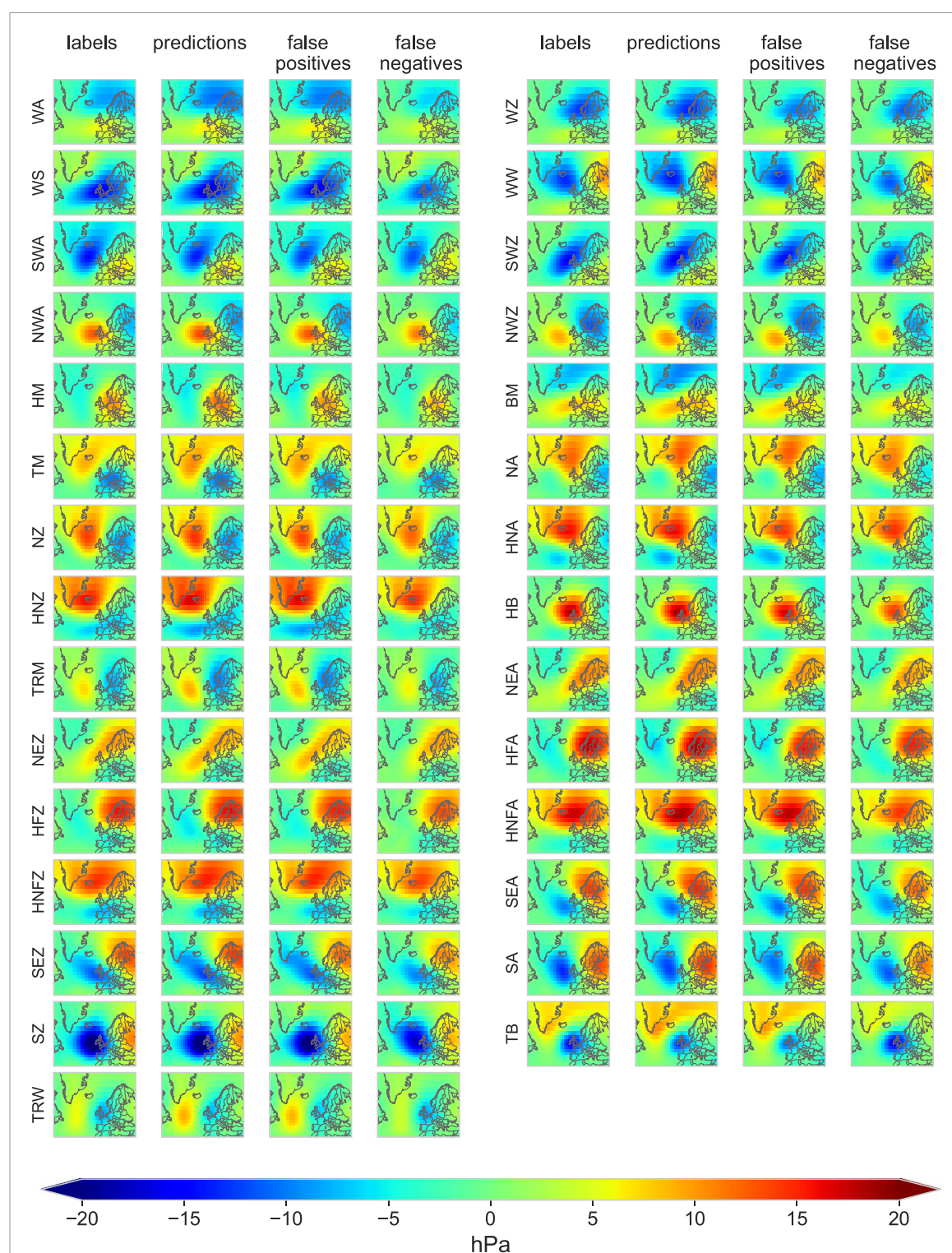


Figure A1. Signature plots of the 29 circulation patterns at slp. Each circulation type (CT) is shown in the respective row for four cases: (column 1) labels: labels showing the indicated CT, (column 2) predictions: deep ensemble predictions showing the indicated CT, (column 3) false positives: signature pattern, when the deep ensemble predicts the indicated CT while labels state differently, (column 4) false negatives: labels stating the indicated CT while deep ensemble predicts differently. The signature plots are derived by calculating the average of all days for which the conditions for this CT are met and subtracting the average of all other CTs. Thus, the composite plots show patterns that distinguish a certain CT from the other types.

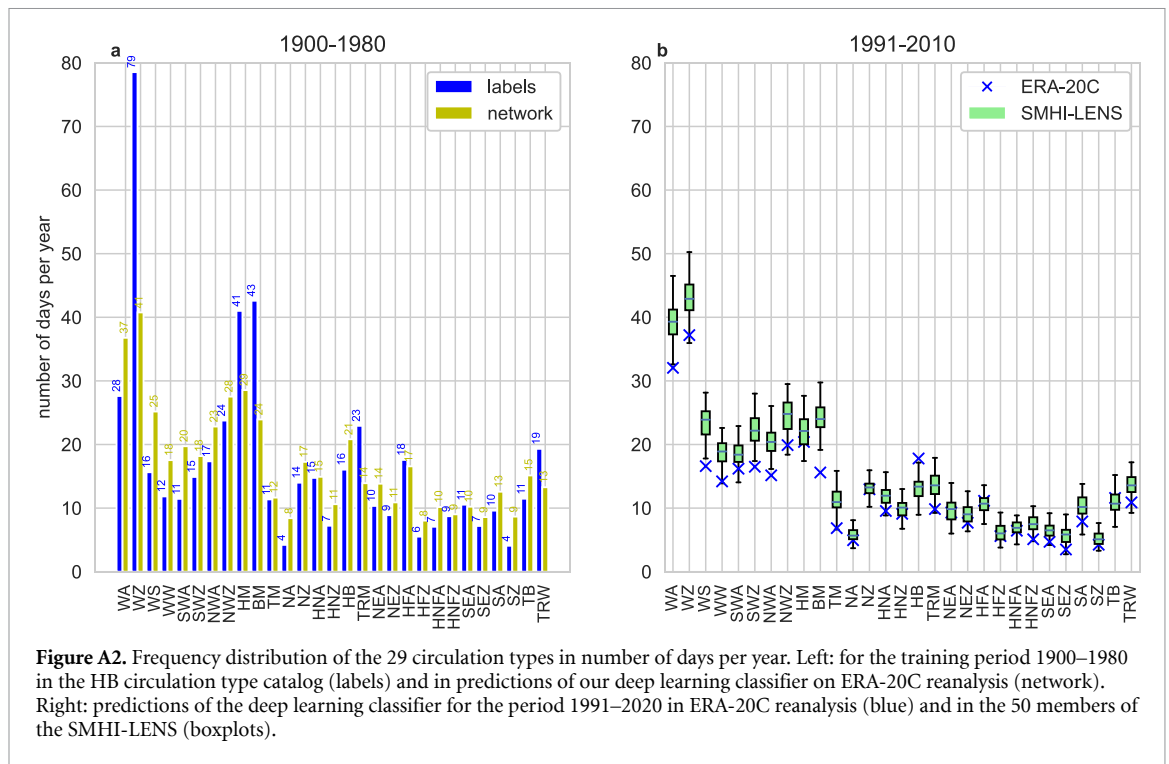
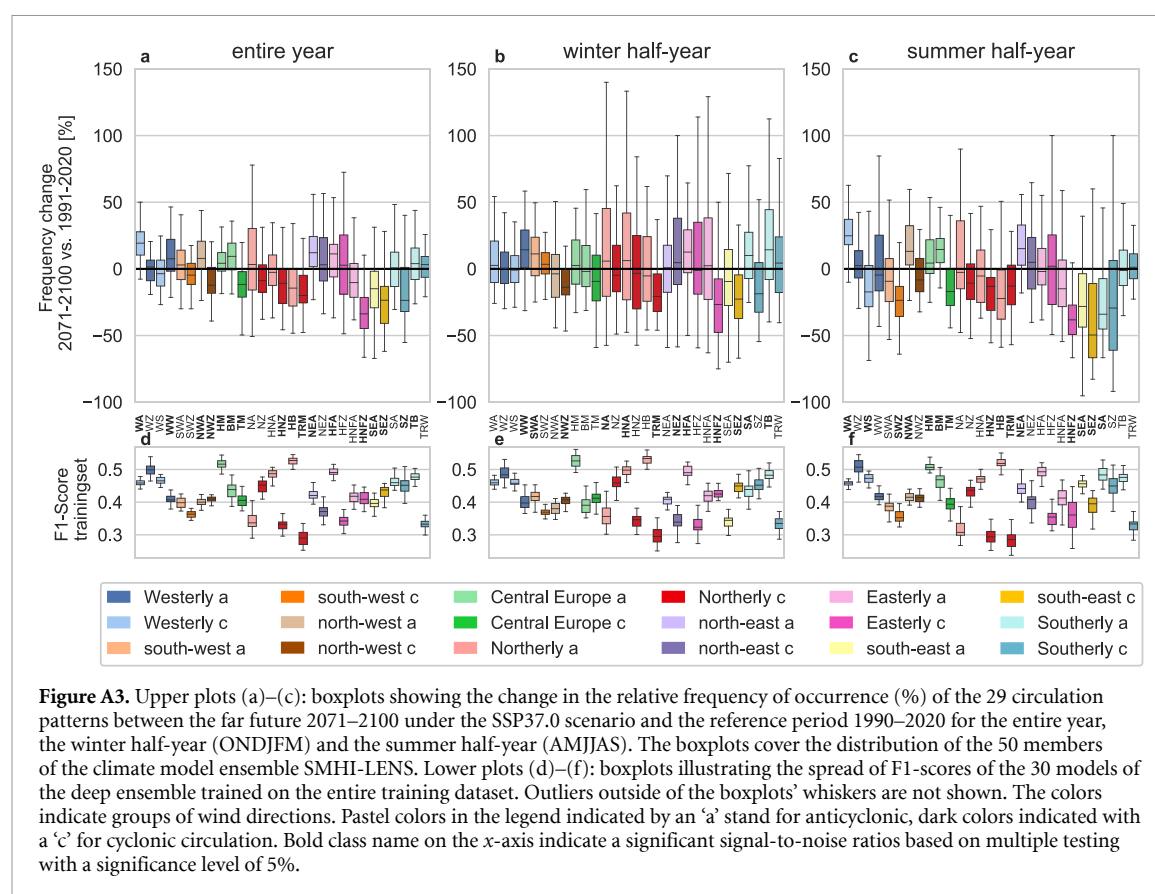


Table A3. S/N-ratios for absolute frequency changes of the 29 circulation types for the total year, winter half-year and summer half-year.

S/N-ratio	WA	WZ	WS	WW	SWA	SWZ	NWA	NWZ	HM	BM	TM	NA	NZ	HNA	HNZ
Total	1.51	0.05	−0.31	0.53	0.15	−0.36	0.6	−0.7	0.31	0.76	−0.65	0.09	−0.36	−0.14	−0.65
Winter half-year	0.23	−0.01	−0.11	0.69	0.53	0.16	−0.11	−0.68	0.13	0.05	−0.33	0.16	−0.03	0.19	−0.11
Summer half-year	1.71	0.09	−0.57	−0.02	−0.39	−1.22	0.8	−0.43	0.34	0.86	−0.67	0.03	−0.39	−0.28	−0.73

S/N-ratio	HB	TRM	NEA	NEZ	HFA	HFZ	HNFA	HNFA	SEA	SEZ	SA	SZ	TB	TRW
Total	−0.71	−0.93	0.68	0.23	0.34	0.01	−0.37	−1.62	−0.67	−1.09	−0.02	−0.66	0.29	0.24
Winter half-year	−0.18	−0.85	0.08	0.22	0.4	0.05	0.06	−0.83	0.29	−0.81	0.42	−0.47	0.55	0.11
Summer half-year	−0.73	−0.5	0.78	0.14	0.04	−0.04	−0.46	−1.71	−0.77	−0.9	−0.78	−0.6	−0.02	0.2



ORCID iDs

M Mittermeier  <https://orcid.org/0000-0002-8668-281X>

M Weigert  <https://orcid.org/0000-0003-4400-134X>

D Rügamer  <https://orcid.org/0000-0002-8772-9202>

References

- Aalbers E E, Lenderink G, van Meijgaard E and van den Hurk B J J M 2018 Local-scale changes in mean and heavy precipitation in western Europe, climate change or internal variability? *Clim. Dyn.* **50** 4745–66
- Beck C, Jacobeit J and Jones P D 2007 Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central Europe since 1780 *Int. J. Climatol.* **27** 473–91
- Benjamini Y and Hochberg Y 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing *J. R. Stat. Soc. B* **57** 289–300
- Cannon A J 2020 Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5–CMIP6 *Environ. Res. Lett.* **15** 064006
- Cawley G C and Talbot N L C 2010 On over-fitting in model selection and subsequent selection bias in performance evaluation *J. Mach. Learn. Res.* **11** 2079–107
- Deser C, Phillips A, Bourdette V and Teng H 2012 Uncertainty in climate change projections: the role of internal variability *Clim. Dyn.* **38** 527–46
- Matschullat J 2021 Synoptic circulation changes over central Europe from 1900 to 2100: reanalyses and coupled model intercomparison project phase 6 *Int. J. Climatol.* **42** 4062–77
- Hess P and Brezowsky H 1952 *Katalog der Großwetterlagen Europas* Ber Dt Wetterd in der USZone 33
- Hoffmann P and Spekat A 2021 Identification of possible dynamical drivers for long-term changes in temperature and rainfall patterns over Europe *Theor. Appl. Climatol.* **143** 177–91
- Hoy A, Hänsel S and Maugeri M 2020 An endless summer: 2018 heat episodes in Europe in the context of secular temperature variability and change *Int. J. Climatol.* **40** 6315–36
- Hoy A, Jaagus J, Sepp M and Matschullat J 2013 Spatial response of two European atmospheric circulation classifications (data 1901–2010) *Theor. Appl. Climatol.* **112** 73–88
- Huguenin M F, Fischer E M, Kotlarski S, Scherrer S C, Schwierz C and Knutti R 2020 Lack of change in the projected frequency and persistence of atmospheric circulation types over central Europe *Geophys. Res. Lett.* **47** e2019GL086132
- Huntingford C, Jeffers E S, Bonsall M B, Christensen H M, Lees T and Yang H 2019 Machine learning and artificial intelligence

- to aid climate change research and preparedness *Environ. Res. Lett.* **14** 124007
- Huth R, Beck C, Philipp A, Demuzere M, Ustrnul Z, Cahynová M, Kysely J and Tveito O E 2008 Classifications of atmospheric circulation patterns: recent advances and applications *Ann. New York Acad. Sci.* **1146** 105–52
- James P M 2006a An assessment of European synoptic variability in Hadley Centre Global Environmental models based on an objective classification of weather regimes *Clim. Dyn.* **27** 215–31
- James P M 2006b An objective classification method for Hess and Brezowsky Grosswetterlagen over Europe *Theor. Appl. Climatol.* **88** 17–42
- Jenkinson A and Collison F 1977 *An Initial Climatology of Gales Over the North Sea* (Bracknell: Meteorological Office. Synoptic Climatology Branch Memorandum No. 62)
- Kingma D P and Ba J 2014 Adam: a method for stochastic optimization (arXiv:1412.6980)
- Krogh A and Vedelsby J 1994 Neural network ensembles, cross validation and active learning *Advances in Neural Information Processing Systems* vol 7
- Kučerová M, Beck C, Philipp A and Huth R 2017 Trends in frequency and persistence of atmospheric circulation types over Europe derived from a multitude of classifications *Int. J. Climatol.* **37** 2502–21
- Kurth T et al 2017 Deep learning at 15pf: Supervised and semi-supervised classification for scientific data (arXiv:1708.05256v1)
- Lakshminarayanan B, Pritzel A and Blundell C 2017 Simple and scalable predictive uncertainty estimation using deep ensembles *Advances in Neural Information Processing Systems* vol 30
- Lewis D D, Schapire R E, Callan J P and Papka R 1996 Training algorithms for linear text classifiers *Proc. 19th Annual International ACM SIGIR Conf. on Research and Development in Information Retrieval* pp 298–306
- Liu Y, Racah E, Prabhat, C, Khosrowshahi A, Lavers D, Kunkel K, Wehner M and Collins W 2016 Application of deep convolutional neural networks for detecting extreme weather in climate datasets (arXiv:1605.01156v1)
- Maher N, Milinski S and Ludwig R 2021 Large ensemble climate model simulations: introduction, overview and future prospects for utilising multiple types of large ensemble *Earth Syst. Dyn.* **12** 401–18
- Minářová J, Müller M, Clappier A, Hänsel S, Hoy A, Matschullat J and Kašpar M 2017 Duration, rarity, affected area and weather types associated with extreme precipitation in the ore mountains (Erzgebirge) region, central Europe *Int. J. Climatol.* **37** 4463–77
- Mittermeier M, Braun M, Hofstätter M, Wang Y and Ludwig R 2019 Detecting climate change effects on Vb cyclones in a 50–member single–model ensemble using machine learning *Geophys. Res. Lett.* **46** 14653–61
- Mittermeier M, Weigert M, Miller J and Banze A 2022 Trained deep ensemble and python code for the classification of atmospheric circulation types over Europe in climate model data (Repository name: Deep learning classification of atmospheric circulation types over Europe) (available at: <https://github.com/mmittermeier/Deep-learning-classification-of-atmospheric-circulation-types-over-Europe>)
- Müller M U, Ekhtiari N, Almeida R M and Rieke C 2020 Super-resolution of multispectral satellite images using convolutional neural networks *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **V-1-2020** 33–40
- Opitz J and Burst S 2021 Macro fl and macro fl (arXiv:1911.03347)
- Petrow T, Zimmer J and Merz B 2009 Changes in the flood hazard in germany through changing frequency and persistence of circulation patterns *Nat. Hazards Earth Syst. Sci.* **9** 1409–23
- Poli P et al 2016 ERA-20c: an atmospheric reanalysis of the twentieth century *J. Clim.* **29** 4083–97
- Racah E, Beckham C, Maharaj T, Kahou S E, Prabhat and Pal C 2016 Extremeweather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events (arXiv:1612.02095v2)
- Ringer M A et al 2006 The physical properties of the atmosphere in the new hadley centre global environmental model (HadGEM1). Part II: aspects of variability and regional climate *J. Clim.* **19** 1302–26
- Rolnick D et al 2019 Tackling climate change with machine learning (arXiv:1906.05433v2)
- Snoek J, Larochelle H and Adams R P 2012 Practical Bayesian optimization of machine learning algorithms *Advances in Neural Information Processing Systems* vol 25
- Sulikowska A and Wypych A 2020 How unusual were June 2019 temperatures in the context of European climatology? *Atmosphere* **11** 697
- Sýkorová P and Huth R 2020 The applicability of the Hess–Brezowsky synoptic classification to the description of climate elements in Europe *Theor. Appl. Climatol.* **142** 1295–309
- Uppala S M et al 2005 The ERA-40 re-analysis *Q. J. R. Meteorol. Soc.* **131** 2961–3012
- Vautard R, Yiou P, Otto F, Stott P, Christidis N, van Oldenborgh G J and Schaller N 2016 Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events *Environ. Res. Lett.* **11** 114009
- Verdecchia M, Visconti G, D’Andrea F and Tibaldi S 1996 A neural network approach for blocking recognition *Geophys. Res. Lett.* **23** 2081–4
- Werner P C and Gerstengarbe F W 2010 Katalog der Großwetterlagen Europas (1881–2009) nach Paul Hess und Helmut Brezowsky *PIK Report Potsdam*. Nr. 119, 7. Aufl
- Woollings T, Hannachi A and Hoskins B 2010 Variability of the North Atlantic eddy-driven jet stream *J. R. Meteorol. Soc.* **136** 856–68
- Wyser K, Koenigk T, Fladrich U, Fuentes-Franco R, Karami M P and Kruschke T 2021 The SMHI large ensemble (SMHI-lens) with EC-Earth3. 3.1 *Geosci. Model Dev.* **14** 4781–96