Climate Dynamics (2024) 62:9927–9949
https://doi.org/10.1007/s00382-024-07432-7

ORIGINAL ARTICLE

Optimizing climate model selection in regional studies using an adaptive weather type based framework: a case study for extreme heat in Belgium

Fien Serras[1] · **Kobe Vandelanotte**[2,3] · **Ruben Borgers**[1] · **Bert Van Schaeybroeck**[2,4] · **Piet Termonia**[2,3] · **Matthias Demuzere**[5] · **Nicole P. M. van Lipzig**[1]

Received: 5 April 2024 / Accepted: 30 August 2024 / Published online: 14 September 2024

Abstract

Selecting climate model projections is a common practice for regional and local studies. This process often relies on local rather than synoptic variables. Even when synoptic weather types are considered, these are not related to the variable or climate impact driver of interest. Therefore, most selection procedures may not sufficiently account for atmospheric dynamics and climate change impact uncertainties. This study outlines a selection methodology that addresses both these shortcomings. Our methodology first optimizes the Lamb Weather Type classification for the variable and region of interest. In the next step, the representation of the historical synoptic dynamics in Global Climate Models (GCMs) is evaluated and accordingly, low-performing models are excluded. In the last step, indices are introduced that quantify the climate change signals related to the impact of interest. Using these indices, a scoring method results in assessing the suitability of GCMs. To illustrate the applicability of the methodology, a case study of extreme heat in Belgium was carried out. This framework offers a comprehensive method for selecting relevant climate projections, applicable in model ensemble-based research for various climate variables and impact drivers.

Keywords CMIP6 · Model selection · Lamb Weather Type classification · Extreme heat

1 Introduction

The latest release of the Intergovernmental Panel on Climate Change's (*IPCC*) Sixth Assessment Report [*AR6*, IPCC (2023)] is heavily based on the sixth generation of the Coupled Model Intercomparison Project (*CMIP*), denoted as CMIP6 (Eyring et al. 2016). Important enhancements were made to the CMIP6 Global Climate Models (GCMs) leading to an overall better performance than previous CMIP generations as documented by e.g. Fan et al. (2020) and Brands (2022a), and a higher global climate sensitivity (Zelinka et al. 2020; De et al. 2022). The IPCC (2023) even states that "robust climate information is increasingly available at regional scales for impact and risk assessments", indicating that GCMs are becoming sufficiently advanced to be used for decision-making on a regional scale. However, the use

✉ Fien Serras
fien.serras@kuleuven.be

Kobe Vandelanotte
kobe.vandelanotte@meteo.be

Ruben Borgers
ruben.borgers@kuleuven.be

Bert Van Schaeybroeck
bert.vanschaeybroeck@meteo.be

Piet Termonia
piet.termonia@meteo.be

Matthias Demuzere
matthias@b-kode.be

Nicole P. M. van Lipzig
nicole.vanlipzig@kuleuven.be

[1] Department of Earth and Environmental Sciences, KU Leuven, 3000 Leuven, Belgium

[2] Department of Meteorological and Climatological Research, Royal Meteorological Institute, 1180 Brussels, Belgium

[3] Department of Physics and Astronomy, Ghent University, 9000 Ghent, Belgium

[4] Department of Geography, Ghent University, 9000 Ghent, Belgium

[5] B-Kode VOF, 9000 Ghent, Belgium

of GCM data for local-scale analyses should be approached with caution (Gualdi et al. 2013).

For the climate of Europe, numerous CMIP6 models have undergone thorough evaluation, as evidenced by comprehensive studies, such as Brunner et al. (2020) and Brands (2022a). The outcomes of these analyses have been included in a selection procedure for dynamic downscaling proposed by the EURO-CORDEX (Coordinated Regional Climate Downscaling Experiment) community (Sobolowski et al. 2023). Their selection criteria include the availability of data for planned analyses, the selection of models with a good overall performance, and the spread of the future climate change signal. While these guidelines and the initially recommended models in Sobolowski et al. (2023) have helped research groups across Europe in their model selection for various analyses, the evaluation of model performance is conducted holistically for a large variety of variables and the entirety of Europe. Therefore, it is not optimized for certain climate impact drivers and different geographical regions. This may play a role as some models may excel in certain regions, such as coastal areas, but may be suboptimal in others such as in mountainous terrains (Brands 2022a; Di Virgilio et al. 2022). Furthermore, some models may be better suited for specific variables (Kotlarski et al. 2014). McSweeney et al. (2015), for instance, has conducted a thorough analysis of surface temperature and precipitation, evaluating models for large regions and examining the potential ranges in key variables.

Understanding the synoptic-scale patterns in climate models is essential, as is shown by the study of Vautard et al. (2023) on heatwaves. They show that, although the CMIP6 models effectively capture how thermodynamics drive temperature rising in Europe, the models often overlook the significant role dynamic processes play in these temperature changes. Moreover, the authors suggest the model dynamics to be responsible for the underestimated trend in extreme temperatures for Western Europe. In other studies, an evaluation method of the synoptic-scale atmospheric dynamics based on weather types is getting more attention and is being studied, e.g. by Otero et al. (2018), Brands (2022a) and Fernández-Granja et al. (2023). Regarding the weather types, a very common method is the Lamb Weather Type classification (LWT) which has already been used in different works, to understand the influence of regional climate on specific variables. For instance, Brisson et al. (2011) demonstrated the impact of weather types on precipitation in coastal regions, characterizing certain weather types as "wet". Another study by Tomczyk and Owczarek (2020) associated the occurrence of heat stress in Poland with the presence of high-pressure systems. Additionally, Hoogeveen and Hoogeveen (2023) developed a temperature-specific weather-typing classification to study the relation between the increasing temperatures and the origin of the airflow.

However, to the best of our knowledge, a method that uses climate impact driver-tailored weather typing for model selection has not been proposed so far. The demonstrated influence of weather types on both the region and variable of interest underscores the need for a methodology that integrates both aspects. A comprehensive framework for the tailored selection of large simulation ensembles that relies on climate projections and is based on synoptic weather types could prove beneficial for a range of purposes, such as analyzing changes in synoptic patterns and downscaling relevant models.

This study proposes a selection methodology that establishes a link between synoptic weather types and variables of interest, enabling the evaluation of models based on pre-established influences. Drawing inspiration from the guidelines proposed by Sobolowski et al. (2023) and the analysis of McSweeney et al. (2015), this study adapts and expands upon these principles to formulate a selection methodology. This approach facilitates the identification of both the most and least extreme model members as well as models with an average signal for a specific region and variable of interest, thereby contributing to an informed decision-making process grounded in relevant climate projections.

2 Materials and methods

2.1 Overview of the selection methodology

To select the models that are best suited for a specific climate change signal or climate impact driver over a region of interest, we developed the methodology outlined in Fig. 1.

The methodology consists of three main steps, each described in the following sections. The previously defined climate change signal (e.g.e. extreme heat) or climate impact driver can be described by one or more variables of interest (e.g. maximum temperature) that will be used in the different steps of the methodology. In the first step (Sect. 2.1.1), the observational data are used to adjust the existing Lamb Weather Type classification to a variable- and region-specific classification. The second step (Sect. 2.1.2) consists of the evaluation of the atmospheric dynamics of the different GCMs and subsequently eliminates low-performing GCMs whose weather types poorly match the observational dataset. The practical considerations of this step revolve around the data availability given the variable and region of interest. The third step (Sect. 2.1.3) classifies the remaining models based on a set of indices related to the climate change signal for the variable of interest. These indices are dependent on the variable of interest. The data availability can again be an important practical consideration. This methodology results in an overview of the climate change signals among the different GCMs. This overview can be combined with practical

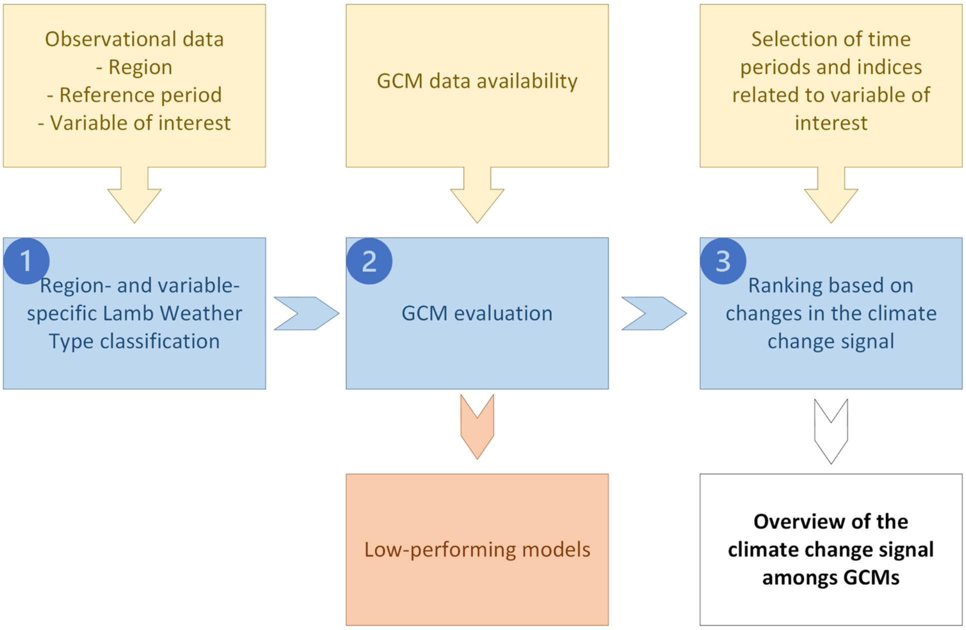


Fig. 1 Overview of the different steps of the selection methodology. The boxes on the top row represent the requirements for each step. The boxes on the second row indicate the different steps of the selection process, with the numbers indicating the respective steps. The bottom row relates to the outcomes of the associated step

considerations for the final decision, e.g. the subset of GCMs for which the driving data is available for a specific Regional Climate Model (RCM).

2.1.1 Region- and variable-specific LWT classification

Following previous studies of e.g. Huth et al. (2008), Demuzere et al. (2009) and Brands (2022a), the evaluation of different GCMs is carried out by comparing the frequency of occurrence of Lamb Weather Types (LWT) of these models with those of a reference dataset over the historical period. One of the main advantages of this classification method is that the different LWTs can be linked to surface variables such as wind speed, rain, and temperature (Jones et al. 1993; Trigo and DaCamara 2000). While several studies (e.g. Jones et al. 1993; Trigo and DaCamara 2000; Brands et al. 2014) have proven that some of these circulation patterns and certain weather phenomena are strongly related, these relations are dependent on the region as well (Jones et al. 1993; Trigo and DaCamara 2000; Brands et al. 2014). For example, Tomczyk and Owczarek (2020) found that persistent and widespread high-pressure systems above Europe often block zonal circulations leading to periods with very strong heat stress. Brunner et al. (2017) have found similar results when connecting atmospheric blocking to temperature extremes in Europe. Additionally, both Trigo and DaCamara (2000) and Brisson et al. (2011) have concluded that certain weather types resulted in significantly wetter days than others, for Portugal and Belgium respectively.

In the first step of our methodology, we therefore propose two adjustments of the existing LWT classification based on the optimization of the grouping of different weather types. The goal of the optimization is both to increase the inter-group separation and to decrease the intra-group variability of the climate variable of interest. By doing so, this classification is tailored to account for a specific signal of a variable for a specific region.

The original 27-type LWT classification is determined based on a 16-point grid of Mean Sea Level Pressure (*MSLP*) which is used to calculate the resultant flow (*F*) and the total shear vorticity (*Z*). The classification uses *F* and *Z* to classify each moment into one of 27 LWTs (Fig. 2): pure cyclonic (C) and anticyclonic (A) circulation, 8 pure directional types (N, NE, E,..., NW) and 16 hybrid types (combinations of either A or C with any of the directional types) and a 27th type (LF), the unclassified records due to days with low flows. Details on this classification can be found in Supp. A. Our suggested adjustments to the original LWT classification are two-fold and result in a 13-type reduced LWT (*rLWT*) classification:

1. Removal of the existing hybrid category: all hybrid weather types are merged with the pure directional types, which is in line with previous studies (e.g. Trigo and DaCamara 2000; Demuzere et al. 2009). The latter study showed that the frequencies of occurrence of the hybrid types are relatively small compared to the other weather types over their region of interest, and additionally, that differences between the pure directional type and its related hybrid types are smaller than differences between directional types. Comparable outcomes were obtained in this study, as illustrated in Supp. B.

2. Introduction of the weak/strong vorticity separation: this step consists of the introduction of four new weather types. These are obtained by splitting the anticyclonic and cyclonic weather types into weak (w) and strong (s) weather types i.e. wA, sA and wC, sC, by optimizing a separation threshold for the total shear vorticity (Z).

We use the Calinski–Harabasz Index [*CHI*, Calinski and Harabasz (1974), also referred to as the Pseudo-F statistic (Beck and Philipp 2010)], as the object function to optimize the grouping of the weather types and the associated climate variable. The Constrained Optimization BY Linear Approximations [*COBYLA*, Powell (1994)] method was used for optimization. The COBYLA method is a simplex-based, direct search method which employs a trust-region strategy. The method approximates the objective function by using linear models near the current point, while also respecting given constraints (Powell 1994).

The *CHI* for K number of weather types in dataset D of the variable of interest for N time instances = $[d_1, d_2, d_3,..., d_N]$ is defined as:

$$CHI = \left[\frac{\sum_{k=1}^{K} n_k \parallel a_k - a \parallel^2}{K-1}\right]\left[\frac{\sum_{k=1}^{K}\sum_{i=1}^{n_k} \parallel d_i - a_k \parallel^2}{N-K}\right]^{-1} \quad (1)$$

where n_k and a_k are the number of time instances and the average for the variable of interest of the k-th weather type, respectively, a is the global average for the variable of interest over all weather types and d_i the value of the variable of interest of data point i within weather type k. The *CHI* measures the variability within each weather type by comparing the variable of interest at different times when that weather type occurs to its average. It also assesses the separation between clusters by comparing the average of each weather type to the global average. Higher *CHI* values indicate that the weather types and related values for the variable of interest are dense and separated. To determine the optimal separation threshold for the total shear vorticity, we iteratively change this threshold, reclassify the weather types and calculate the *CHI*. Generally, there are no clear cut-off values for the *CHI*; therefore, Calinski and Harabasz (1974) suggest that the value which gives a clear peak or an abrupt increase/decrease in the *CHI* plot should be chosen. These values were used as different initial guesses for COBYLA in order to avoid obtaining local maxima. This procedure is illustrated in supplementary Fig. 14.

2.1.2 CMIP6 evaluation

The developed regional and variable-specific rLWT classification is used to evaluate the different CMIP6 models in representing synoptic dynamics. We have calculated the model frequencies of each of the 13 rLWTs and compared them with the corresponding LWT frequencies derived from ERA5 for the reference period. The Perkins Skill Score [*PSS*, Perkins et al. (2007)] was used as a metric to evaluate the correspondence between these frequency distributions, defined as:

$$PSS = \sum_k \min(f_{M,k}, f_{O,k}), \quad (2)$$

with k the weather type, $f_{M,k}$ the frequency of weather type k in the model distribution and $f_{O,k}$ the frequency of weather type k in the reference distribution. A perfect match between both distributions would lead to a *PSS* of 1, whereas a complete mismatch leads to a *PSS* of 0. Following an empirical statistical rule, all models with *PSS* values below the *PSS* average minus the *PSS* standard deviation are omitted from further analysis.

2.1.3 Scoring based on climate change signal

The second and last step of this methodology concerns the selection of models based on their climate change signal for a future period of interest. This requires the definition of different indices related to the climate change signal of interest, e.g. HUMIDEX in case of humid heat or return periods in case of extreme precipitation.

We capture the relationship between the weather types and the variable of interest in a newly developed index, by assigning weights to certain percentiles for the variable of interest and investigating the change in weather types. As such we address whether the occurrence of several weather types, relevant to the variable of interest, is projected to change. More specifically, we introduce a new index called the “Weather Type Change Index” (*WTCI*) to quantify the future change in the frequency of weather types relevant to the variable of interest (*var*). The following formulas are used for the calculation:

$$WTCI = \sum_{k=1}^{K} \Delta f_k \cdot \lambda_k, \quad (3)$$

$$\lambda_k = \frac{\alpha_k}{\sum_{k=1}^{K} \alpha_k}, \quad (4)$$

$$\alpha_k = \begin{cases} f_{hist}(var < x \mid k) \text{ if } x < 50 \\ f_{hist}(var \geq x \mid k) \text{ if } x \geq 50 \end{cases} \quad (5)$$

where Δf_k refers to the frequency change between the future and the historical period. The frequency of historical days (f_{hist}) with weather type k where the variable of interest stays below (< 50) or exceeds (≥ 50) the x-th percentile is used as a weighting factor for each weather type (α_k), which is normalized for all weather types, resulting in λ_k.

The differentiation at $x = 50$ is made to make the *WTCI* generally applicable, e.g. minimum versus maximum temperatures. The percentile can be extreme (e.g. 1% or 99%) but could target the median of the distribution (50%) as well. Additionally, this is illustrated by the case study that will be discussed in Sect. 2.3.2.

Once all indices are determined, an overall score for all models can be calculated as follows:

1. Determine the minimum and maximum value for each index over each future period.
2. Scale the indices using minimum–maximum scaling.
3. Average the scaled indices for each model. This gives the overall score for each model.

This strategy results in an overall ranking for the models included, given a climate change signal or climate impact driver. Upon using multiple indices, their dependence should be investigated as one may wish to account for the relation between indices. An example of such analysis would be to perform a principal component analysis.

2.2 Data

2.2.1 ERA5

The fifth generation atmospheric reanalysis (*ERA5*) of the global climate of the European Centre for Medium-Range Weather Forecasts (*ECMWF*) covers the period from January 1940 to the present and is available at a resolution of about 0.25 ° (Hersbach et al. 2023).

For this study, we used the ERA5 data between 1985 and 2014, following the end date of most CMIP6 historical runs as well as the guidelines of the World Meteorological Organization (2017) to use 30-year periods. Three variables were used: mean sea level pressure (*MSLP*), 2 m temperature (*T*) and 2 m dewpoint temperature (*Tdew*). Additionally, following the recommendations of the ECMWF, daily minimum and maximum temperatures (*Tmin* and *Tmax*, respectively) were obtained from the hourly 2 m temperature data (ECMWF 2020). Relative humidity (*RH2m*) was obtained by combining *T* and *Tdew*. The calculations for the *RH2m* are included in Supp. D. The ERA5 data was bilinearly interpolated to a common 1 ° × 1 ° grid, following e.g. Broderick and Fealy (2015) and Kim et al. (2020).

2.2.2 CMIP6

An ensemble of CMIP6 models with one member per model is considered. A summary of the CMIP6 models can be found in Table 1. For each model, the historical and the Shared Socioeconomic Pathways (*SSP*, limited to SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5) were retained if all required variables were available at a daily frequency ("Retained" in Table 1). Following Brunner et al. (2020) and Brands (2022b), it is recommended to consider different model families to incorporate the widest possible range of model components, hence the model families are also included in the table. A limited number of models that the IPCC used in the AR6 report (IPCC 2023), were not included in the whole analysis due to data unavailability (e.g. AWI-CM-1-1-MR and CIESM), or due to known errors in the most recently-published data [e.g. CESM2, CESM2-WACCM, information obtained from ES-DOC (2016)]. The same variables as for ERA5 were extracted, excluding the dewpoint temperature but including the relative humidity, and we adopted the same abbreviations. The *RH2m* was preferred over the *Tdew* as this was available for more models. Similarly to ERA5, all data was bilinearly interpolated to a common 1 ° × 1 ° grid.

2.3 Methodology applied to a case study: extreme heat in Belgium

This section uses the developed selection methodology to answer the question: "Which five CMIP6 models have the most extreme summer temperature signals under different GWLs for Belgium?". Given that Belgium is a relatively small country, we considered the coordinate 50°N–5°E to be representative of Belgium. Note that this is a very specific research question to illustrate the application of the methodology and that the framework is flexible to be used for different CMIP6 selection criteria.

We choose to define model periods based on different Global Warming Levels (GWLs). The method of using GWLs instead of transient scenarios has become a common practice in the scientific literature as well as in communication with society and stakeholders (IPCC 2023). Several studies (e.g. Seneviratne et al. 2016; Tebaldi and Knutti 2018; Li et al. 2020a) have found that the climate change signal at different GWLs is almost entirely independent of scenario and time horizon for most variables. Additionally, they stated that the signal at a GWL is nearly linearly related to regional climate effects. GWLs represent the change in the global mean surface temperature throughout the fixed duration, which is usually a 20- or 30-year period (IPCC 2023; World Meteorological Organization 2017), from a future period concerning a reference period (here 1985–2014) where the average global temperatures from 1850 until 1900 are used as a baseline. This baseline was obtained by averaging three datasets: NOAAGlobalTemp (Zhang et al. 2023), HadCRUT5 (Morice et al. 2021) and Berkely-Earth (Rohde and Hausfather 2020). In the

Table 1 Overview of the CMIP6 model runs used in this work

CMIP6 model	Run	Model family	Retained	SSP	References
ACCESS-CM2	r1i1p1f1	HadGAM/UM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Bi et al. (2020)
ACCESS-ESM1-5	r1i1p1f1	HadGAM/UM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Ziehn et al. (2020)
BCC-CSM2-MR	r2i1p1f1	CAM			Wu et al. (2019)
BCC-ESM1	r1i1p1f1	CAM			Wu et al. (2020)
CanESM5	r1i1p1f1	CanAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Swart et al. (2019)
CESM2	r4i1p1f1	CAM			Danabasoglu et al. (2020)
CESM2-FV2	r1i1p1f1	CAM			Danabasoglu et al. (2020)
CESM2-WACCM	r1i1p1f1	CAM			Danabasoglu et al. (2020)
CESM2-WACCM-FV2	r1i1p1f1	CAM			Danabasoglu et al. (2020)
CMCC-CM2-HR4	r1i1p1f1	CAM			Cherchi et al. (2019)
CMCC-CM2-SR5	r1i1p1f1	CAM			Cherchi et al. (2019)
CMCC-ESM2	r1i1p1f1	*CAM*	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Lovato et al. (2022)
CNRM-CM6-1	r1i1p1f2	ARPEGE	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Voldoire et al. (2019)
CNRM-CM6-1-HR	r1i1p1f2	ARPEGE	✓	SSP5-8.5	Voldoire et al. (2019)
CNRM-ESM2-1	r1i1p1f2	ARPEGE	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Séférian et al. (2019)
EC-Earth3	r1i1p1f1	IFS	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Döscher et al. (2022)
EC-Earth3-AerChem	r1i1p1f1	IFS	✓	SSP3-7.0	Döscher et al. (2022)
EC-Earth3-CC	r1i1p1f1	IFS	✓	SSP2-4.5, SSP5-8.5	Döscher et al. (2022)
EC-Earth3-Veg	r1i1p1f1	IFS	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Döscher et al. (2022)
EC-Earth3-Veg-LR	r1i1p1f1	IFS	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Döscher et al. (2022)
FGOALS-f3-L	r1i1p1f1	*GAMIL*			He et al. (2020)
FGOALS-g3	r1i1p1f1	*GAMIL*	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Li et al. (2020b)
GFDL-CM4	r1i1p1f1	GFDL-AM	✓	SSP2-4.5, SSP5-8.5	Held et al. (2019)
GFDL-ESM4	r1i1p1f1	GFDL-AM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Dunne et al. (2020)
GISS-E2-2-G	r1i1p1f1	GISS-E2			Rind et al. (2020)
HadGEM3-GC31-LL	r1i1p1f3	HadGAM/UM	✓	SSP1-2.6, SSP2-4.5, s SSP5-8.5	Roberts et al. (2019)
HadGEM3-GC31-MM	r1i1p1f3	HadGAM/UM	✓	SSP1-2.6, SSP5-8.5	Roberts et al. (2019)
IITM-ESM	r1i1p1f1	*GFS*	✓	SSP1-2.6	Swapna et al. (2015)
INM-CM4-8	r1i1p1f1	INM-AM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Volodin et al. (2018)
INM-CM5-0	r1i1p1f1	INM-AM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Volodin et al. (2017)
IPSL-CM5A2-INCA	r1i1p1f1	LMDZ			Dufresne et al. (2013)
IPSL-CM6A-LR	r1i1p1f1	LMDZ	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Boucher et al. (2020)
IPSL-CM6A-LR-INCA	r1i1p1f1	LMDZ			Dufresne et al. (2013)
KACE-1-0-G	r2i1p1f1	HadGAM/UM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0	Lee et al. (2020a)
KIOST-ESM	r1i1p1f1	GFDL-AM	✓	SSP2-4.5, SSP5-8.5	Pak et al. (2021)
MIROC6	r1i1p1f1	MIROC-AGCM/CCSR	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Tatebe et al. (2019)
MIROC-ES2L	r1i1p1f2	MIROC-AGCM/CCSR	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Hajima et al. (2020)
MPI-ESM1-2-HR	r1i1p1f1	ECHAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Mauritsen et al. (2019)
MPI-ESM1-2-LR	r1i1p1f1	ECHAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Mauritsen et al. (2019)
MPI-ESM-1-2-HAM	r1i1p1f1	ECHAM	✓	SSP3-7.0	Müller et al. (2018)
MRI-ESM2-0	r1i1p1f1	GSMUV/MRI-AGCM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Yukimoto et al. (2019)
NESM3	r1i1p1f1	ECHAM			Cao et al. (2018)
NorCPM1	r1i1p1f1	CAM			Bethke et al. (2021)
NorESM2-LM	r1i1p1f1	CAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Seland et al. (2020)
NorESM2-MM	r1i1p1f1	CAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0	Seland et al. (2020)
SAM0-UNICON	r1i1p1f1	CAM			Park et al. (2019)
TaiESM1	r1i1p1f1	CAM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Lee et al. (2020b)
UKESM1-0-LL	r1i1p1f2	HadGAM/UM	✓	SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5	Sellar et al. (2019)

"Model family" follows the grouping made by Brands (2022b) and Brunner et al. (2020), with italics referring to GCMs that did not fulfil the grouping criteria. The models that were retained for the case study are indicated in the "Retained" and "SSP" columns following sections, the “(model) period” refers to a future 30-year period around a GWL.

In this work, similar to Sanderson et al. (2011), four different GWLs have been chosen to define model periods: 1.5 °C, 2 °C, 3 °C and 4 °C. The method of Vautard et al. (2014) was followed to determine the periods when the different models reached these GWLs.

To equalize the comparison between the different GCMs, all CMIP6 temperature variables and RH2m were bias-corrected. The equidistant cumulative distribution function matching technique was used, a quantile-mapping approach that explicitly accounts for distribution changes between the projection and historical periods (Li et al. 2010).

2.3.1 Applied temperature-dependent classification

The first step of the selection methodology (as explained in Sect. 2.1.1) consists of the adjustment of the LWT classification. For this, we used the daily *Tmax* of ERA5. The *CHI* reached an optimum (109.81) for a total shear vorticity of 18 (17.78) hPa, corresponding to an optimal separation between weak and strong (anti)cyclones as described in Sect. 2.1.3. An overview of the new, temperature-dependent 13-type classification is given in Fig. 2.

In Fig. 2, it can be seen that the pure directional category has been extended compared to the original classification. Additionally, the separation between weak and strong (anti-)cyclones becomes visible.

2.3.2 Indices for ranking

As mentioned in the methodology, the ranking that will result in the model periods with the highest increase in extreme heat is based on different indices. The selection methodology here is based on three different indices: one related to the weather types, as suggested, and two related to heat. As such, we attempt to capture different characteristics influencing the extreme heat climate change signal. We used two different heat stress indices: the Heat Wave Degree Days (*HWDD*), which relates to the length and intensity of heatwaves, and the humidity index (*HUMIDEX*), which relates to thermal comfort. The three indices address heat on different temporal scales: the *HUMIDEX* is calculated daily, the *WTCI* is calculated with daily data but aggregated over the total period, while the *HWDD* only considers heat waves and is therefore linearly related to the amount of heat. For both heat indices, the change in P95 between the historical period and the future period was calculated and used in this study. Details in the *HWDD* and the *HUMIDEX* can be found in Supp. E.

Figure 3 shows how the different rLWTs relate to the daily maximum temperature. Continental air masses (E, SE and S) generally result in warmer days, while oceanic air masses (W, SW) result in colder temperatures. These findings result in the third index for the ranking: the *WTCI*, as introduced in Eq. (5). Here, we use *Tmax* for *var* (variable of interest) and *x* equals 90, reflecting our focus on extreme heat. Positive values indicate an increase in weather types related to extreme temperatures. As weather

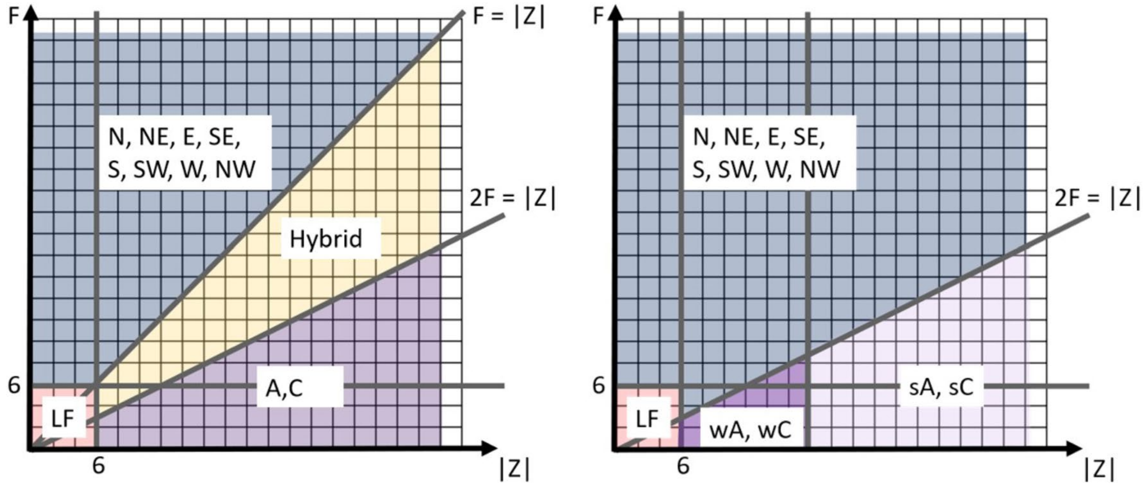
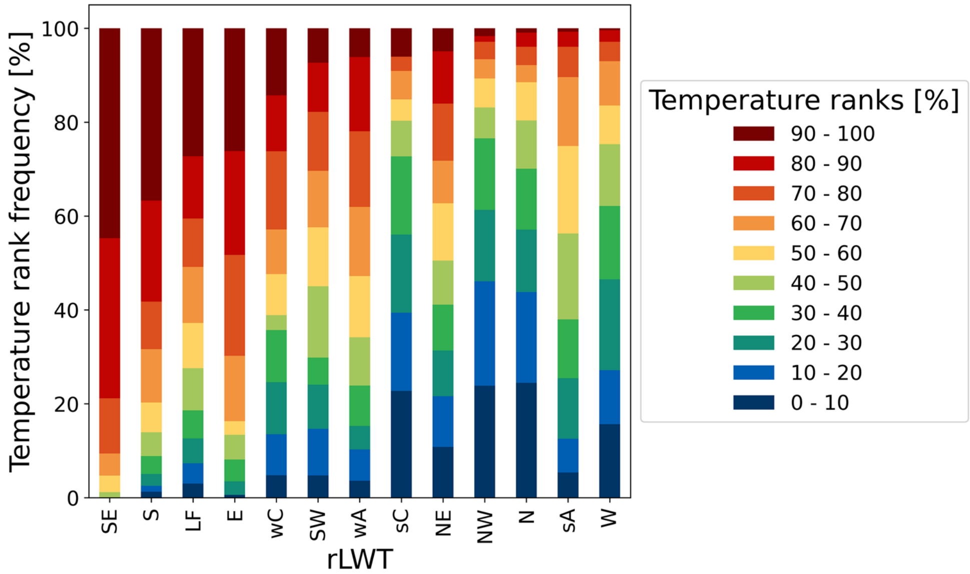


Fig. 2 Graphic overview of the Lamb Weather Type classification based on the total shear vorticity (*Z*, x-axis) and the resultant flow (*F*, y-axis). Both flow and vorticity units are geostrophic and are expressed in hPa per 10 ° latitude. The left graph refers to the original LWT classification, as introduced by Jenkinson and Collison (1977). The right graph represents the adjusted classification for extreme heat in Belgium. “LF” stands for the Low Flow/Unclassified weather type. “N, NE,..., NW” represent the different wind directions. “A” and “C” stand for the (Anti)cyclonic weather types, while the hybrid weather types are a combination of either wind direction with either (anti) cyclonic weather type. Prefixes “w” and “s” stand for “weak” and “strong”, respectively

Fig. 3 Relation between the weather types of the reduced Lamb Weather Type classification and the corresponding maximum temperatures for JJA over the period 1985–2014. The weather types on the x-axis are ordered on decreasing frequency of each weather type above the 90th temperature percentile



types are also related to wind direction and humidity, this index indirectly accounts for changes in atmospheric conditions.

3 Results

3.1 CMIP6 evaluation

We assess the evaluation of the atmospheric dynamics of the different CMIP6 models. Figure 4 shows rLWT summer (JJA) frequencies where the panels per model are sorted by decreasing *PSS* (Eq. 2). Four out of 32 models fall below the performance threshold of 0.815, which was defined as the

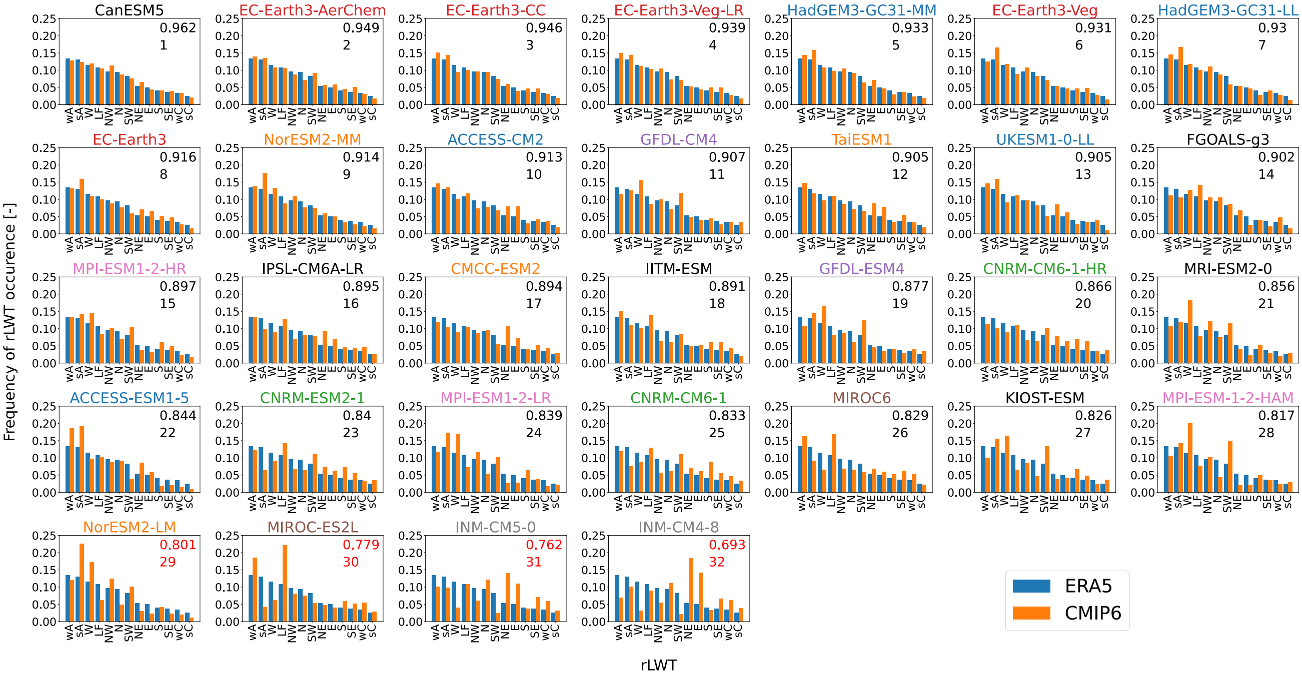


Fig. 4 Barplots representing the frequency of the different reduced Lamb Weather Types (rLWT) for different CMIP6 models in JJA. The Perkins Skill Score (top number) as well as its rank compared to the other models (bottom number) is shown. Red values indicate a PSS performance worse than $\mu - \sigma$ (here: 0.815). The colours of the model names refer to the different model families as given in Table 1

average *PSS* minus the standard deviation of the *PSS* of the included GCMs. These models are deemed to represent the large-scale atmospheric circulations inferior to other models and are consequently excluded from further analysis. In general, lower *PSS* values can be observed when a model strongly over-/underestimates the frequency of a weather type that occurs either very often (e.g. sA for MIROC-ES2L) or very little (e.g. E for INM-CM4-8).

The worst-performing models (in red in Fig. 4) have clear differences in the frequencies of the different rLWTs compared to ERA5. Both INM-CM4-8 and INM-CM5-0 overestimate the Northeasterly (NE) and Easterly (E) weather types while underestimating the Southwesterly (SW) and Westerly (W) weather types. MIROC-ES2L strongly overestimates the Low Flow (LF) and weak Anticyclonic (wA) weather types, while underestimating the strong Anticyclonic (sA) weather type. This might be partially explained by the model underestimating the vorticity, as this leads to more wA and less sA. And lastly, NorESM2-LM mainly overestimates sA and W and underestimates the LF and Northerly (N) weather types. In general, two model families are less capable of representing the atmospheric dynamics: INM-AM (average *PSS*: 0.727) and MIROC (average *PSS*: 0.804).

The 10 best models predominantly originate from two model families (Table 1): the HadGAM/UM (HadGEM3-GC31-LL, HADGEM3-GC31-MM and ACCESS-CM2), and all models from the EC-Earth Consortium. The rLWT frequencies of each of these eight models have a high correspondence (lowest *PSS* = 0.913) with those of ERA5, indicating that the models represent the atmospheric dynamics very well during the JJA of the reference period.

3.2 Model selection based on the change of the climate change signal

After the evaluation and subsequent elimination of the CMIP6 models, we now rank models for Belgium based on three indices that are related to extreme heat. The results for these indices for the different GWL periods are shown in Fig. 5.

For the selected indices there is a big spread among the different models and the GWLs. Additionally, this spread increases with increasing GWL, indicating that a warmer world will on average lead to more heat stress but also to more uncertainty on the heat-stress signal. Overall, the median *WTCI* increases only slightly with increasing GWL.

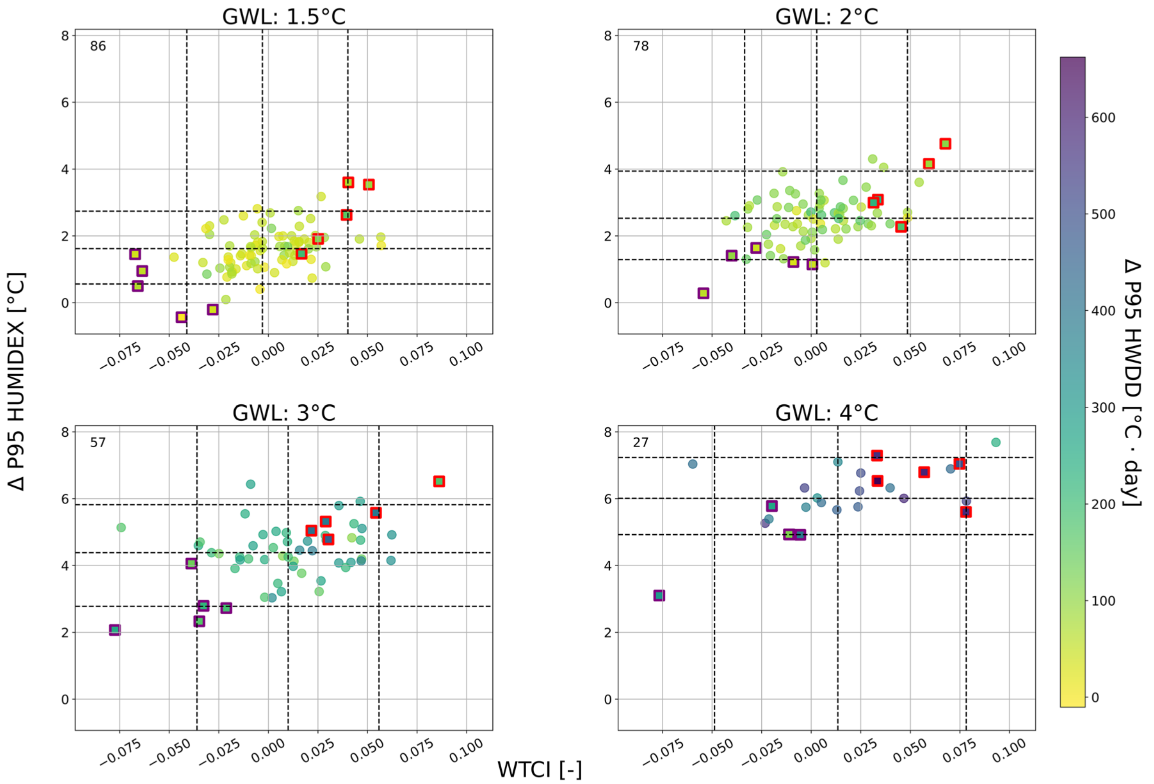


Fig. 5 Graphical overview of the positioning of the different CMIP6 models for three different indices (x-axis, y-axis and colorbar) and four GWLs (different subplots). Red squares indicate the top five models in the final overview, purple squares indicate the bottom five models. The number of models that reach a certain GWL is indicated in the top left corner of each plot. The dashed lines represent the P5, P50 and P95 of the *WTCI* and Δ P95 *HUMIDEX* for each GWL

Additionally, the P95 increases with GWL whereas the P5 remains similar over all GWLs. Lastly, both the P95 and P5 strongly change between 3 °C and 4 °C, with a larger range for 4 °C. In contrast, the change in the *HUMIDEX* consistently increases with GWL, similar to the change in *HWDD*. In general, it can be observed that both the *HWDD* and *HUMIDEX* are more sensitive to thermodynamic changes than to dynamic changes, which are encapsulated in the *WTCI*. These findings are in line with the results of Otero et al. (2018) who stated that changes in dynamics have a smaller influence on variations in European temperatures compared to changes in thermodynamics. Therefore, the *WTCI* seems to be rather less sensitive to the GWL compared to the *HUMIDEX* and the *HWDD*.

The presented methodology results in a ranking of the models for each GWL and for different SSPs by combining the scaled indices. As the correlation between the three chosen indices was overall relatively low in our study (0.82, 0.41 and 0.30), we did not include any weighting of ranks.

The five models with the highest (red squares in Fig. 5) and the lowest scores (purple squares) in the ranking for extreme heat are mostly situated in the first quadrant and third quadrant respectively. Quadrants are here based on the median values for the *WTCI* and the change in the 95th percentile of the *HUMIDEX*. This visually demonstrates the efficacy of the developed selection methodology. Models that have a high (low) overall score also have a high (low) score for the different indices. Lastly, the answer to our case-study-specific question, the five models with the highest score and thus the overall most extreme heat signal for Belgium are given in Table 2.

4 Discussion

4.1 CMIP6 evaluation

The model evaluation is predicated on the assumption of Sobolowski et al. (2023), stating that "realistic models will produce more realistic future projections, because they can represent processes correctly". Following this, we aimed to develop a methodology that could limit the risk of selecting models with errors in their synoptic-scale model dynamics. By introducing the performance threshold in the CMIP6 evaluation, we excluded models that are less capable of representing the atmospheric dynamics of the region of interest. However, as stated by Knutti et al. (2010), the question of whether a model is good or bad often depends on the intended application. As a consequence, we developed the variable-specific LWT classification and applied this for heat in Belgium. By doing so, we ensured that the model evaluation was based on heat-related weather types. As errors in atmospheric dynamics are often inherited in RCMs after downscaling, our developed methodology can for example be used to limit this transfer of errors in a physically consistent way.

One example of the application potential of the selection methodology is downscaling GCMs with specific climate change signals, and more specifically, the propagation of the climate change signal when downscaling (Liang et al. 2008). By introducing weather types, that act on the synoptic scale, we aimed at developing a method that would limit the risks of an unwanted propagation, e.g. the RCM has a strong cooling signal while the GCM has a strong warming signal.

The general evaluation of the CMIP6 models as previously discussed is in line with the results of Brands (2022a) for the northern hemisphere and the ones of the CORDEX white paper [('Bra21LambEUR' in the summary table, Sobolowski et al. (2023)], where the MAE was used as an evaluation criterion. We found that our correlation between *PSS* and MAE was equal to − 1, indicating that when applied to the same categorical data, the *PSS* and MAE are opposites of each other. The Spearman correlation between our *PSS* and the MAE of the results of (Brands 2022a) for the northern hemisphere is − 0.88 for models included in both studies and − 0.87 when compared with Europe [(Bra21LambEUR, Sobolowski et al. (2023)]. The negative sign of the correlation indicates that models that adequately represent the atmospheric dynamics feature high *PSS* values and low MAE values. The correlation is remarkably high despite the differences in the reference period (1979–2005 versus JJA of

Table 2 Models with the highest score for the most extreme change in heat signal over Belgium, with the reference period of 1985-2014

1.5 °C	2 °C	3 °C	4 °C
ACCESS-ESM1-5 SSP5-8.5 (0.812)	MRI-ESM2-0 SSP5-8.5 (0.808)	CanESM5 SSP3-7.0 (0.803)	IPSL-CM6A-LR SSP5-8.5 (0.825)
MRI-ESM2-0 SSP5-8.5 (0.812)	ACCESS-ESM1-5 SSP5-8.5 (0.782)	ACCESS-ESM1-5 SSP5-8.5 (0.756)	CanESM5 SSP3-7.0 (0.817)
MRI-ESM2-0 SSP3-7.0 (0.751)	TaiESM1 SSP5-8.5 (0.755)	MRI-ESM2-0 SSP5-8.5 (0.731)	EC-Earth3-Veg SSP5-8.5 (0.799)
TaiESM1 SSP3-7.0 (0.716)	MRI-ESM2-0 SSP5-8.5 (0.707)	KIOST-ESM SSP5-8.5 (0.726)	EC-Earth3 SSP5-8.5 (0.796)
EC-Earth3-Veg SSP2-4.5 (0.636)	NorESM2-MM SSP2-4.5 (0.660)	EC-Earth3-Veg SSP5-8.5 (0.717)	TaiESM1 SSP5-8.5 (0.771)

1985–2014) and reference data (ERA-Interim and JRA-55 versus ERA5). Additionally, the MAE as given by Brands (2022a) is the median value of all calculations over the mid-to-high latitudes in the Northern Hemisphere, while the values of this study are limited to the calculations representative of Belgium only. Similar to their findings, the models of the Earth Consortium all had a high *PSS*. The affinity of these models to ERA5 has previously been partly related to the fact that they both are based on the ECMWF IFS and thus share similar model components (Hersbach et al. 2020; Döscher et al. 2022; Brands 2022b). ? also found that a high correlation is present between the northern and southern hemisphere. Given the similarity between these results, it indicates that the Lamb Weather Type approach for GCM evaluation is consistent over different regions.

In general, we can conclude that the overall assessment as mentioned in the CORDEX white paper (Sobolowski et al. 2023) provides a comprehensive overview of the performance of CMIP6 models over Europe. Our framework can be used complementary, as it offers intricate details for specific regions and periods which could contribute to the selection process.

4.2 Selection methodology

Our selection methodology consists of three steps. In each included step, we aimed to develop a method that considers the uncertainties in the large-scale atmospheric dynamics and accounts for the climate change signal of interest. Admittedly, we recognize the sensitivity of our developed methodology to the number of included models and the length of the considered periods. Both are discussed below.

To start, our evaluation is limited to 32 models as the remaining CMIP6 models lacked the data required for the further steps of this study. A more extensive analysis, including 48 CMIP6 models, can be found in Supp. F. Executing the evaluation with other models inherently leads to a different performance threshold for *PSS*, in this case 0.807. However, the same models were excluded, which seems to indicate that the method is robust and rather independent of the number of included models.

Furthermore, to investigate the sensitivity of the method to the length of the model periods, we applied our selection methodology to 20-year periods (reference period: 1995–2014), instead of the 30-year periods used so far. Here, the same threshold of 18 hPa was found for the optimal separation of weak and strong (anti-)cyclonic weather types. This might indicate that such separation could be applied in similar research where the maximum temperatures and related weather types in summer are studied.

Figure 6 shows that the *PSS*s of the 20-year periods (y-axis) are similar to those obtained using the 30-year periods (x-axis) and have a correlation of 0.979. This shows that the evaluation with the *PSS* is rather robust. This robustness can be partially related to the fact that the *PSS* results in one value for each model and will have relatively similar results over different periods.

Figure 7 depicts the differences in scoring for each GWL. It becomes clear that there is a rather large impact of the period length on the final scoring based on the three indices for the climate change signals, despite moderate to high correlations between both period lengths.

As can be seen, the scoring varies most for models at the centre of each graph ([0.40, 0.60] in Fig. 7). This larger variation can be related to the values of the indices; the indices of the models in the middle will be more similar compared to the most and the least extreme models. As a consequence, minor differences between both periods can lead to a substantial change in score for the models in the middle. The top and bottom scores, on the other hand, have changed less, suggesting that the length of the periods has a relatively small effect on the most extreme models. Therefore, our selection methodology can be considered most suitable for selecting the models with the highest and the lowest change in their climate signal.

Our case study was intended to find five CMIP6 models that have the highest extreme heat signal. Table 3 shows the findings with the reference period of 1995-2014 and period lengths of 20 years. Some models and SSPs are the same as in Table 2, but there are differences. When comparing values in both tables, it becomes clear that the scoring is very sensitive to the considered periods as the scores of the same models also differ. This emphasizes that our selection methodology has succeeded in selecting case-study-specific models of interest.

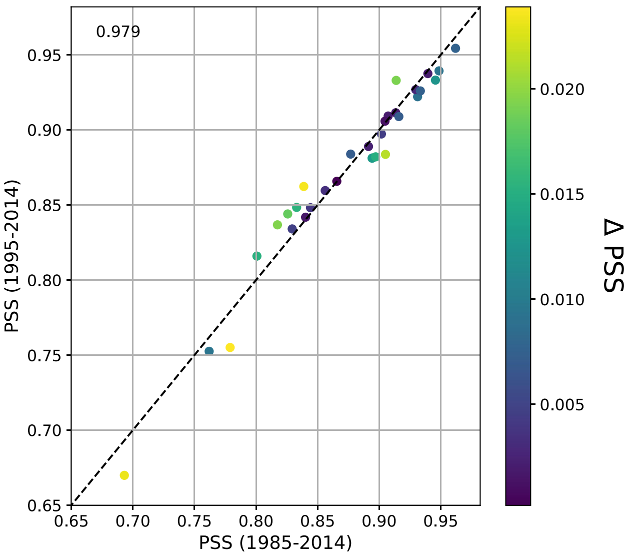


Fig. 6 *PSS* for a 30-year reference period (1985–2014) compared to the *PSS* for a 20-year reference period (1995–2014). The correlation is indicated in the top left corner

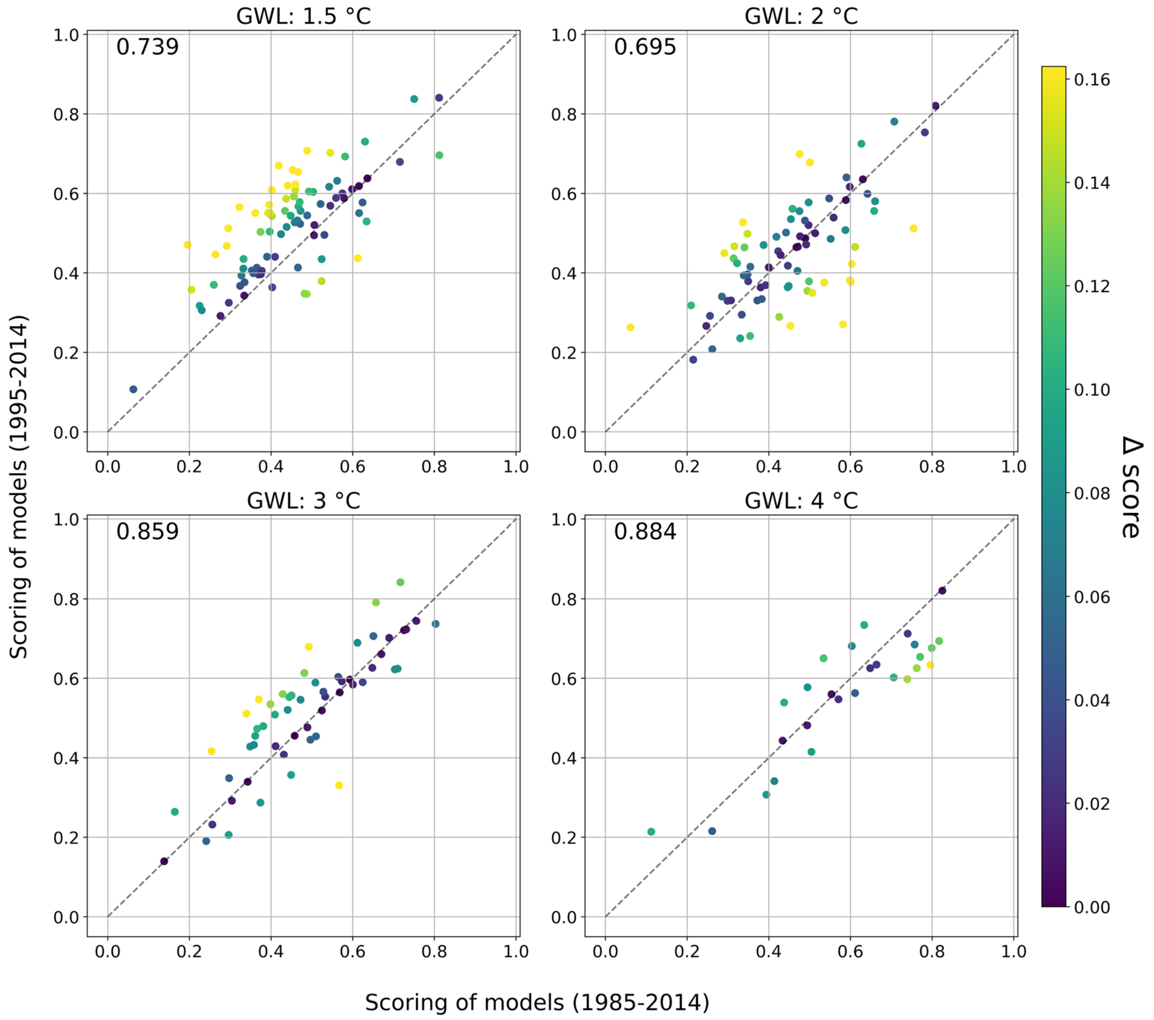


Fig. 7 Scores of the CMIP6 models obtained using 30-year periods (reference period: 1985–2014) against their scores obtained using 20-year periods (reference period: 1995–2014) for GWL 1.5 °C, 2 °C, 3 °C and 4 °C. The correlations are indicated in the top left corner

In addition to the sensitivity of the method for the period length, we also investigated its sensitivity to the consistency of models used. More specifically, we performed the same analysis but only using those models for which all warming levels were available, i.e. those that reach a 4 °C warming (within the 21st century), instead of all models available. Figure 8 and Table 4 are similar to Fig. 5 and Table 2, but only include the models that reach all GWLs. Additionally, Fig. 9 shows the P5, P50 and P95 for the initial analysis as

Table 3 Same as Table 2, but with the reference period of 1995-2014

1.5 °C	2 °C	3 °C	4 °C
MRI-ESM2-0 SSP5-8.5 (0.841)	MRI-ESM2-0 SSP5-8.5 (0.820)	EC-Earth3-Veg SSP5-8.5 (0.842)	IPSL-CM6A-LR SSP5-8.5 (0.820)
MRI-ESM2-0 SSP3-7.0 (0.837)	MRI-ESM2-0 SSP3-7.0 (0.781)	IPSL-CM6A-LR SSP5-8.5 (0.791)	UKESM1-0-LL SSP5-8.5 (0.734)
MRI-ESM2-0 SSP1-2.6 (0.730)	ACCESS-ESM1-5 SSP5-8.5 (0.753)	ACCESS-ESM1-5 SSP5-8.5 (0.744)	MRI-ESM2-0 SSP5-8.5 (0.712)
UKESM1-0-LL SSP2-4.5 (0.707)	EC-Earth3 SSP1-2.6 (0.725)	CanESM5 SSP3-7.0 (0.737)	ACCESS-ESM1-5 SSP3-7.0 (0.701)
GFDL-ESM4 SSP1-2.6 (0.702	CNRM-CM6-1 SSP1-2.6 (0.699)	MRI-ESM2-0 SSP5-8.5 (0.723)	CanEMS5 SSP3-7.0 (0.694)

given in Fig. 5 and for the analysis with 27 models from Fig. 8.

In general, it can be observed in both Fig. 8 and Table 4 that the results are similar for the most extreme models. This indicates that our methodology can select the most extreme models and that adding more models has a limited effect on the final ranking. As some of the top five models in Table 2 do not reach 4 °C of warming, they are not included in Table 4. Removing these models affected the minimum–maximum scaling as well, leading to different final scores. Figure 8 shows the influence on the spread as compared to Fig. 5. Overall, the median, P5 and P95 remain similar for the *HUMIDEX* and the *HWDD*, which can also be observed in Fig. 9. Additionally, the median for the *WTCI* remains similar as well, while the P5 and P95 differ more. This can be explained by the fact that, as pointed out in Sect. 3.2, the *WTCI* is rather insensitive to the GWL. This means that removing the less climate-sensitive models has a non-related effect on the *WTCI* signal.

To finalize, we recognize the limitations of our methodology. A limitation of the use of the LWT classification for the evaluation lies in its consideration of a substantial geographical expanse for the calculations, while only yielding a singular output. As such, some information is inherently lost. Additionally, the use of

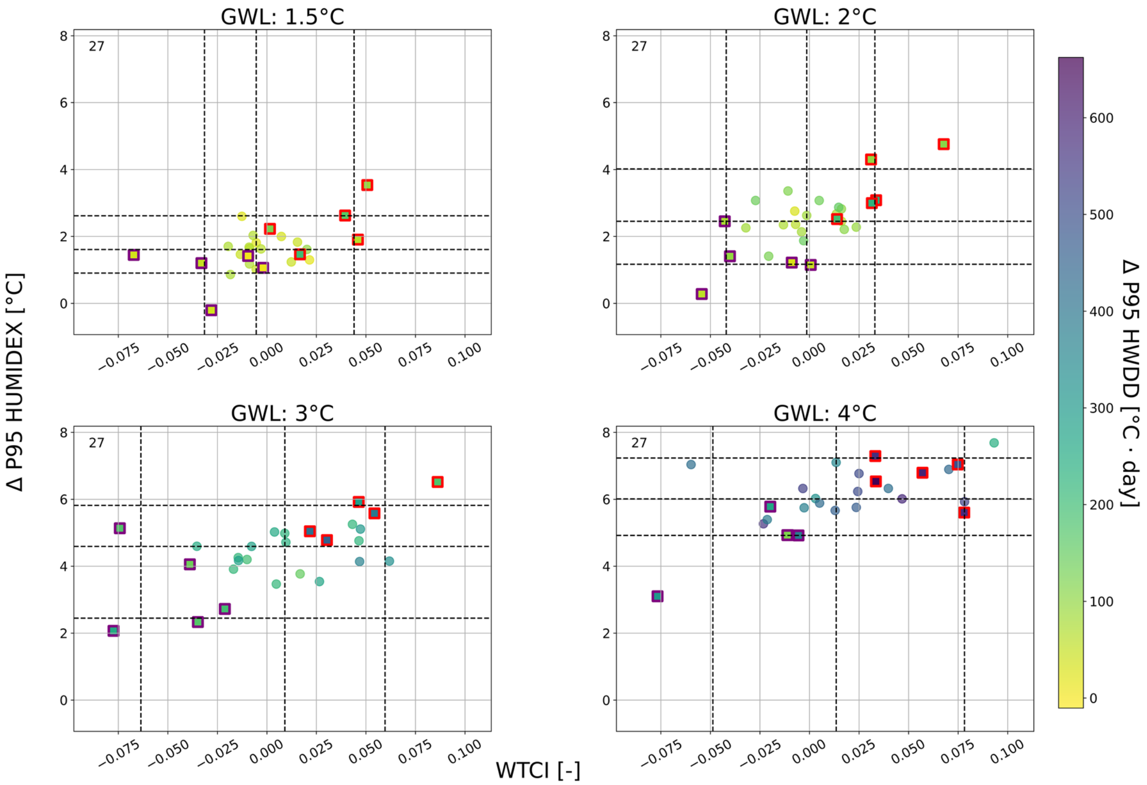


Fig. 8 Same as Fig. 5, but only the 27 models that reach 4 °C are included

Table 4 Same as Table 4, but only the 27 models that reach 4 °C are included

1.5 °C	2 °C	3 °C	4 °C
ACCESS-ESM1-5 SSP5-8.5 (0.822)	MRI-ESM2-0 SSP5-8.5 (0.795)	CanESM5 SSP3-7.0 (0.797)	IPSL-CM6A-LR SSP5-8.5 (0.825)
MRI-ESM2-0 SSP5-8.5 (0.821)	ACCESS-ESM1-5 SSP5-8.5 (0.782)	ACCESS-ESM1-5 SSP5-8.5 (0.756)	CanESM5 SSP3-7.0 (0.817)
TaiESM1 SSP3-7.0 (0.719)	TaiESM1 SSP5-8.5 (0.754)	EC-Earth3-Veg SSP5-8.5 (0.713)	EC-Earth3-Veg SSP5-8.5 (0.799)
CNRM-CM6-1 SSP5-8.5 (0.612)	EC-Earth3 SSP5-8.5 (0.644)	MRI-ESM2-0 SSP5-8.5 (0.706)	EC-Earth3 SSP5-8.5 (0.796)
CNRM-ESM2-1 SSP5-8.5 (0.606)	TaiESM1 SSP3-7.0 (0.606)	EC-Earth3 SSP5-8.5 (0.695)	TaiESM1 SSP5-8.5 (0.771)

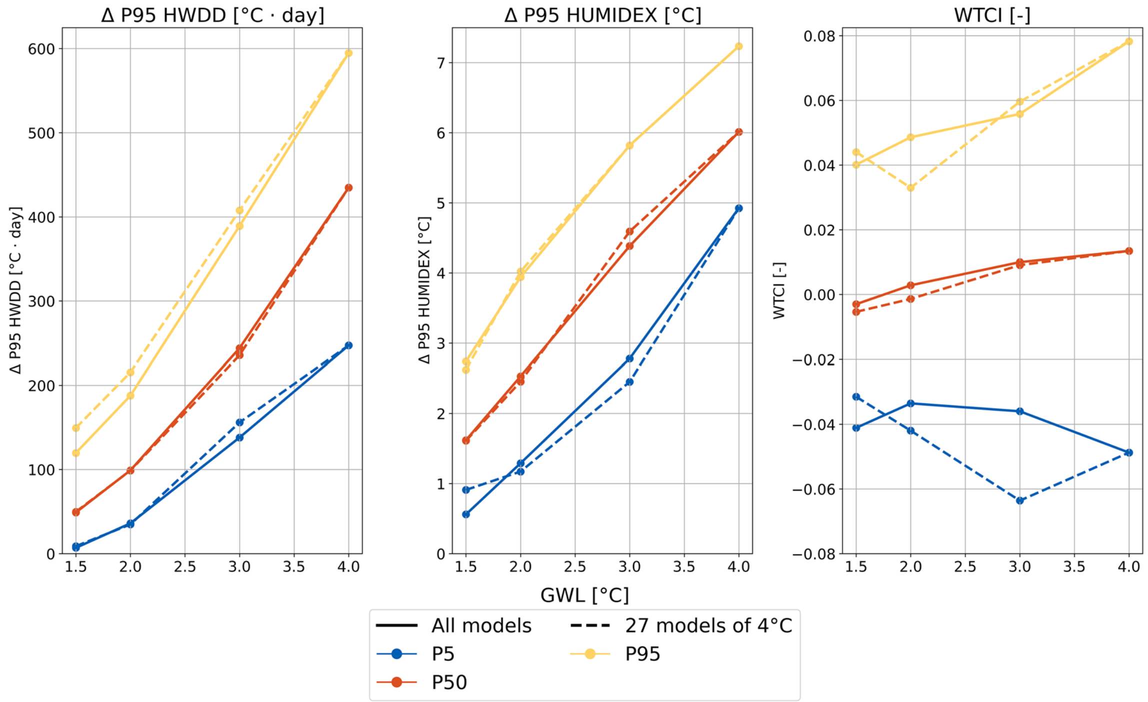


Fig. 9 Comparison between P5 (blue), P50 (red) and P95 (yellow) for all available models (solid line) and the 27 models that reach 4 °C (dashed line)

the LWT classification is limited to mid-to-high latitude regions (Jones et al. 2013). However, a recent study by Fernández-Granja et al. (2023) stated that it can be reliably applied over most areas within the 23.5 ° to 80 ° latitudinal range. And lastly, we did not apply the readjustment or selection methodology to larger and more diverse geographical regions, or other variables of interest, such as precipitation. Thus, a thorough review becomes necessary before applying the methodology to other cases. As a consequence, we cannot ensure that performing the readjustment on other variables would give a similarly strong differentiation for the different weather types. For example, Brisson et al. (2011) showed both the regional and seasonal differences in correlation between geostrophic winds, which they related to LWTs and precipitation. However, the adjustments are optional changes and can be left out in case of more complex variables of interest, such as snow events.

5 Conclusion

This study presents a novel methodology to select climate models for a specific region and variable of interest. The first step includes the adjustment of the existing LWT classification to make it both variable and region-specific and thus, application-specific. The adjusted classification is consequently included in the second step, the evaluation of the atmospheric dynamics through synoptic-scale weather patterns. This ensures that the selection procedure limits the risk of inheriting the major errors in the synoptic-scale model dynamics. The second step of the methodology encompasses a specific analysis of different indices that quantify the climate change signal for the climate impact driver of interest. We developed a new index, the *WTCI*, that quantifies the contribution of a change in weather types to the change in the variable of interest, which could be a specific percentile of temperature, precipitation, or wind ...between two time periods. The *WTCI* can be used together with other indices for a given climate impact driver, like heat stress indices. The GCMs are scored with different indices, which results in an assessment of the suitability of the models for the required application.

We applied the developed methodology for the case study of extreme heat over Belgium on the CMIP6 ensemble using different GWLs. We excluded four low-performing models after model evaluation. As our case study focused on extreme heat, we tailored the *WTCI* and included a index that targets heat comfort (*HUMIDEX*) and one that targets heatwaves (*HWDD*). The *WTCI* was found to give complementary information to these existing heat indices. The CMIP6 evaluation was quite robust concerning changes in the length of the climatological period (20 or 30 years), while the scoring is more sensitive.

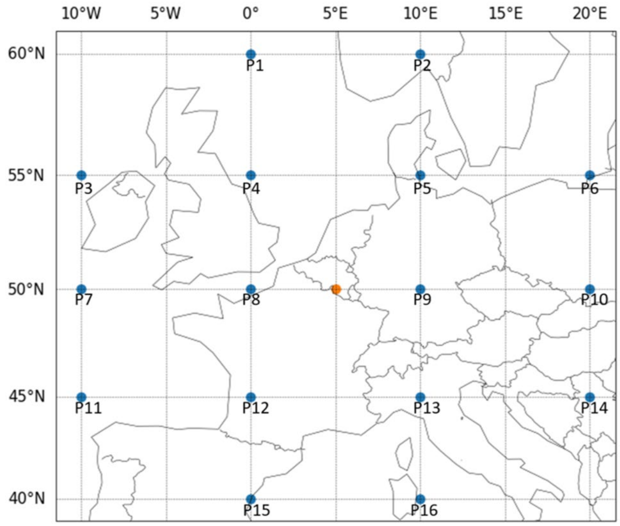


Fig. 10 The 16-point grid as used in the Lamb Weather Type classification. The grid is centred over the orange point within Belgium

We highlight the potential of this method as a general framework for tailored model selection. Our developed selection methodology can contribute to informed decisions when selecting CMIP6 models for different purposes, such as the detailed analysis of climate signals and downscaling of extreme weather events. The framework can be used for selecting model periods to downscale with RCMs. Additionally, the adjusted variable dependent classification can serve as a starting point to analyze the relation between circulation patterns and other climatological variables.

Appendix 1: Lamb Weather Type classification

The Lamb Weather Type classification was developed by Lamb (1950, 1972) as a large-scale circulation classification based on maps of mean sea level pressure (MSLP) for the regional weather over the British Isles. Jenkinson and Collison (1977) translated this empirical classification into an objective classification, known as both the LWT classification and the "Jenksinson–Collison Weather Type" classification. The LWT classification uses the daily Mean Sea Level Pressure *MSLP* (in hPa) at a 16-point grid (Fig. 10) over a spatial extent of 30° longitudes by 20° latitudes to classify the synoptic pattern at its centre point, located at latitude ϕ. Each of the 16 points is spaced 5 ° in latitude (Δlat) and 10 ° in longitude toward the adjacent points. Both the flow units as well as the units for the geostrophic vorticity are expressed in hPa.

The following rules are used to define the LWTs, based on the formulas shown in Table 5:

1. The direction of the flow (in degrees) is given by $\tan^{-1}(W/S) + \epsilon$, where $\epsilon = 180\,°$ if W is positive and $\epsilon = 0$ otherwise. The corresponding wind direction is computed using an eight-direction compass, allowing 45 ° per sector. For example: the Northeastern weather type occurs for wind direction between 22.5 ° and 67.5 °.
2. If |Z| > 2F, then the pattern is strongly cyclonic (Z > 0) or anticyclonic (Z < 0), corresponding to the pure cyclonic and anticyclonic types.
3. If |Z| is between F and 2F, the flow is considered to be of a hybrid type and is characterized by both direction and circulation, leading to sixteen different types.
4. If |Z| < F, the flow is considered straight and corresponds to a Lamb pure directional type. Eight different types exist: N, NE, E, SE, S, SW, W and NW.
5. If both |Z| and F are smaller than the LF-threshold of 6 hPa, as defined by Jenkinson and Collison (1977), the

Table 5 Equations for the different parameters required for the Lamb Weather Type classification, for the Northern Hemisphere

Parameter	Symbol	Equations
Westerly Flow	W	$= \left[(p_{12} + p_{13}) - (p_4 + p_5)\right]/2$
Southerly Flow	S	$= \alpha\left[(p_5 + 2p_9 + p_{13}) - (p_4 + 2p_8 + p_{12})\right]/4$
Resultant Flow	F	$= (S^2 + W^2)^{1/2}$
Westerly Shear Vorticity	WZ	$= \beta\left[\left((p_{15} + p_{16}) - (p_8 + p_9)\right) - \gamma\left((p_8 + p_9) - (p_1 + p_2)\right)\right]/2$
Southerly Shear Vorticity	SZ	$= \frac{\delta}{4}\left[(p_6 + 2p_{10} + p_{14}) - (p_5 + 2p_9 + p_{13}) - (p_4 + 2p_8 + p_{12}) + (p_3 + 2p_7 + p_{11})\right]$
Total Shear Vorticity	Z	$= WZ + SZ$

Where $\alpha = \cos^{-1}(\phi)$, $\beta = \sin(\phi)\sin^{-1}(\phi - \Delta lat)$, $\gamma = \sin(\phi)\sin^{-1}(\phi + \Delta lat)$ and $\delta = cos^{-2}(\phi)$

The p's represent the values of the MSLP (in hPa) as given in Fig. 10, the ϕ is the latitude of the central point and Δlat is the lateral distance among the points

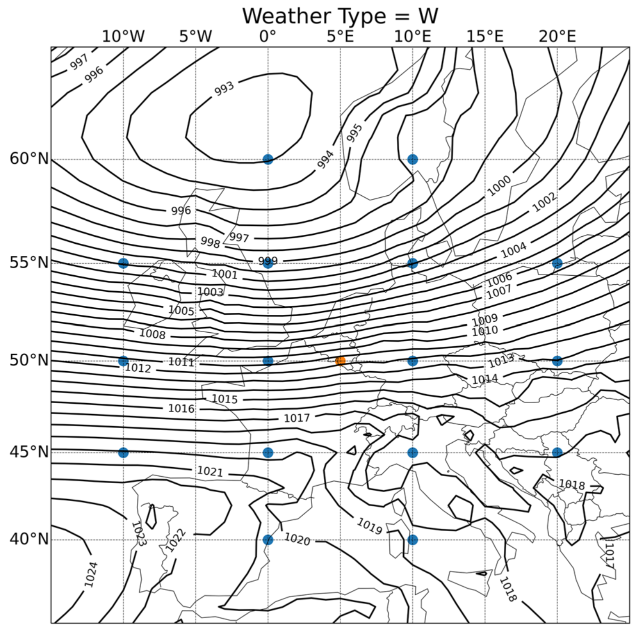


Fig. 11 Averaged MSLP patterns for the W-weather type over the period 1985–2014

flow is indeterminate, corresponding to the Low Flow type (also called the "Unclassified type"). This corresponds to the so-called "barometric swamps" (Grimalt et al. 2013).

Appendix 2: Overview of directional and hybrid resemblance

The figures below show all weather types associated with the western flow, as determined for ERA5. The relative frequency of each weather type over the period 1985–2014 is given in Table 6 (Figs. 11, 12, 13).

Appendix 3: Weak-strong optimization with CHI

Figure 14 below showcases the initial step of the optimization. Here, the *CHI* was calculated for all Z thresholds between 0 hPa and 60 hPa, in order to obtain guesses for the optimization. The *CHI*-values between 0 and 6 and between 39 and 60 are not relevant for the analysis, as the targeted 13 weather types are not present here.

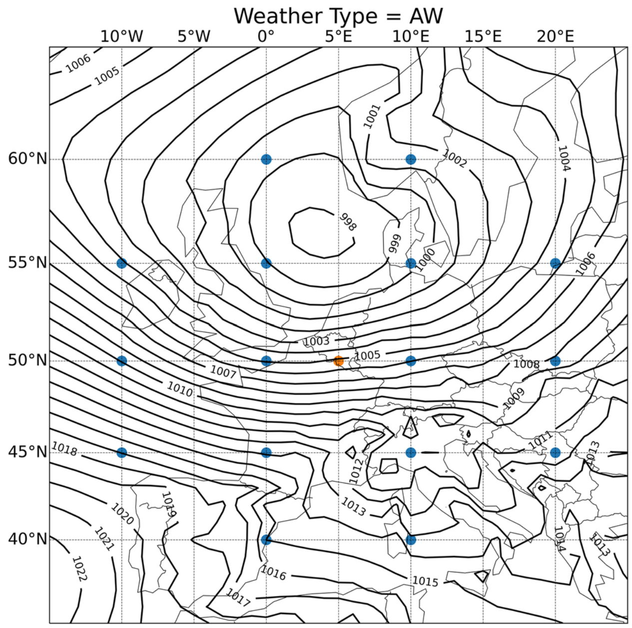


Fig. 12 Averaged MSLP patterns for the AW-weather type over the period 1985–2014

Appendix 4: ERA5: calculation of relative humidity at the surface level

ERA5 lacks direct relative humidity data at the 2-meter level. Therefore, we derived this information by calculating it from the corresponding temperature and dew-point temperature

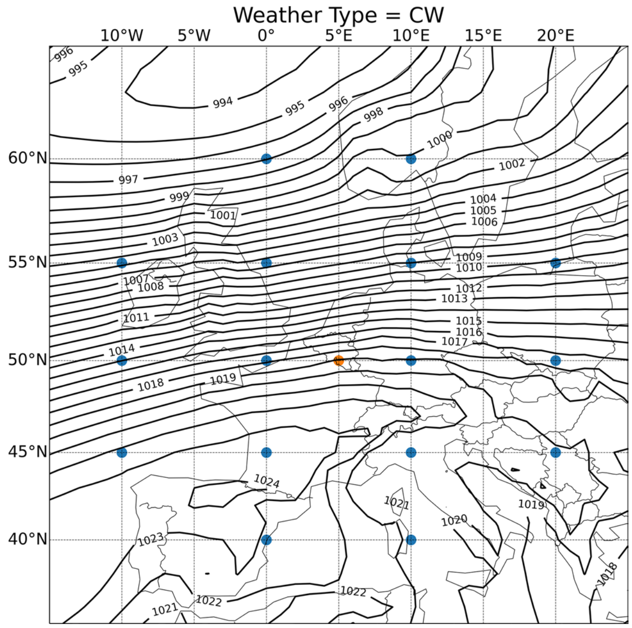


Fig. 13 Averaged MSLP patterns for the CW-weather type over the period 1985–2014

Table 6 Relative frequency (%) of the directional and hybrid weather types over the summer periods of 1985–2014

Direction	Pure	Hybrid cyclonic	Hybrid anticyclonic
N	6.37	4.78	0.80
NE	5.91	3.33	1.16
E	3.95	1.85	0.43
SE	1.78	0.51	0.80
S	1.67	0.58	0.62
SW	4.71	1.27	0.94
W	5.25	2.40	1.12
NW	5.80	2.40	0.62

The LF (10.94%), anticyclonic (23.12%) and cyclonic (6.92%) weather type are excluded from the table

at 2 ms. The calculations follow the steps as described in ECMWF (2016):

$$e_{sat,x,w} = e_{S0} \cdot \exp\left(17.502\left(\frac{x - T_0}{x - 32.19}\right)\right) \tag{6}$$

$$e_{sat,x,i} = e_{S0} \cdot \exp\left(22.587\left(\frac{x - T_0}{x + 0.7}\right)\right) \tag{7}$$

$$e_{sat,x} = \alpha e_{sat,x,w} + (1 - \alpha) e_{sat,x,i} \tag{8}$$

where x can be both the air temperature and the dewpoint temperature at 2 ms, T_0 equals 273.16K and e_{S0} represent the saturated vapour at T_0 and equals 611.21 Pa. and α represents the mixed phase ratio and is calculated as follows:

with

$$\alpha = 0 \quad T <= T_{ice} \tag{9}$$

$$\alpha = {\frac{T - T_{ice}}{T_0 - T_{ice}}}^{2} \quad T_{ice} < T < T_0 \tag{10}$$

$$\alpha = 1 \quad T >= T_{ice} \tag{11}$$

where T_{ice} equals 250.16K.

The relative humidity (%) is:

$$RH = 100 \cdot \frac{e_{sat,Tdew}}{e_{sat,T}} \tag{12}$$

Appendix 5: Heat indices

HWDD

The *HWDD* is a yearly heat-stress indicator specifically developed for Belgium (Brouwers et al. 2015; Wouters et al. 2017) and quantifies the accumulated temperature exceedance during heat-wave days:

$$HWDD = \sum_{i}^{N} \left[\left(Tmin - Tmin_{P90,hist} \right)^{+} + \left(Tmax - Tmax_{P90,hist} \right)^{+} \right] h_i, \tag{13}$$

with $Tmax_{P90,hist}$ and $Tmin_{P90,hist}$ referring to the 90th percentile of Tmax (°C) and Tmin (°C), respectively, of the JJA data over the historical period of 1985–2014. The summation i is over N days and h_i indicates (0 or 1) whether the day i is a heat-wave day i.e. when the 3-day average of the minimum and maximum temperature exceeds their respective thresholds and thus incorporating the temporal persistence of the heat wave. The superscript + refers to the positive difference between the considered maximum/minimum temperature and the respective threshold and equals zero if the difference is negative. By including the minimum temperatures, *HWDD* also takes into account high nocturnal temperatures which can limit the ability of the body to recover during heatwave periods and which has been related to increased mortality among elderly people (Basu 2002; Laaidi et al. 2012).

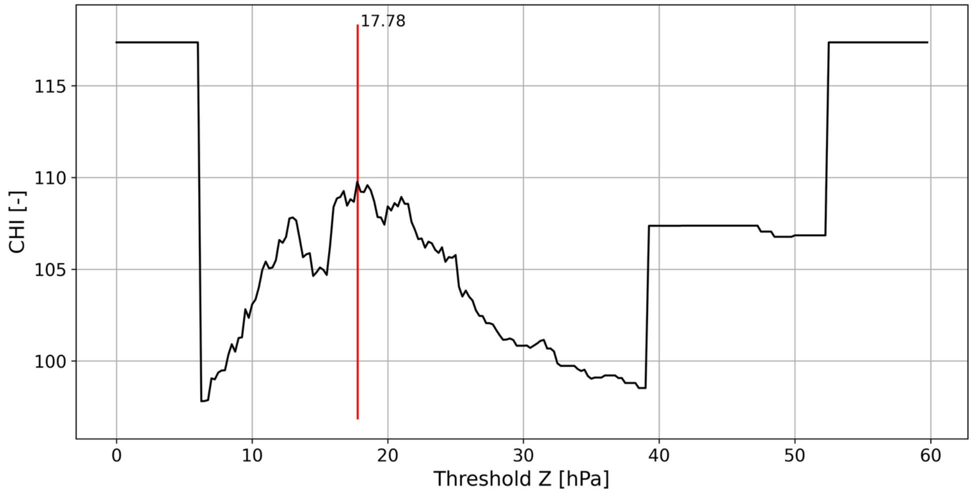
HUMIDEX

The *HUMIDEX* is calculated based on Basara et al. (2010), who adapted the original formula from Masterton and Richardson (1979):

$$\text{HUMIDEX} = T + \frac{5}{9}(e - 10), \quad \text{where} \quad e = 6.112 \cdot 10^{7.5 \cdot T/(237.7+T)} \cdot \frac{RH}{100}, \tag{14}$$

where e refers to the partial vapour pressure (hPa). By incorporating both temperature and humidity, the *HUMIDEX* includes two of the most important variables for addressing thermal (dis)comfort (Epstein and Moran 2006; Fischer et al. 2012). However, we also acknowledge that other indices can be used to describe heat (dis)comfort and that other variables impact heat (dis)comfort, e.g. radiation (Van De Walle et al. 2022).

Fig. 14 CHI as a function of total shear vorticity values ranging from 0 to 60 hPa (top panel). Corresponding CHI step changes are shown in the bottom panel



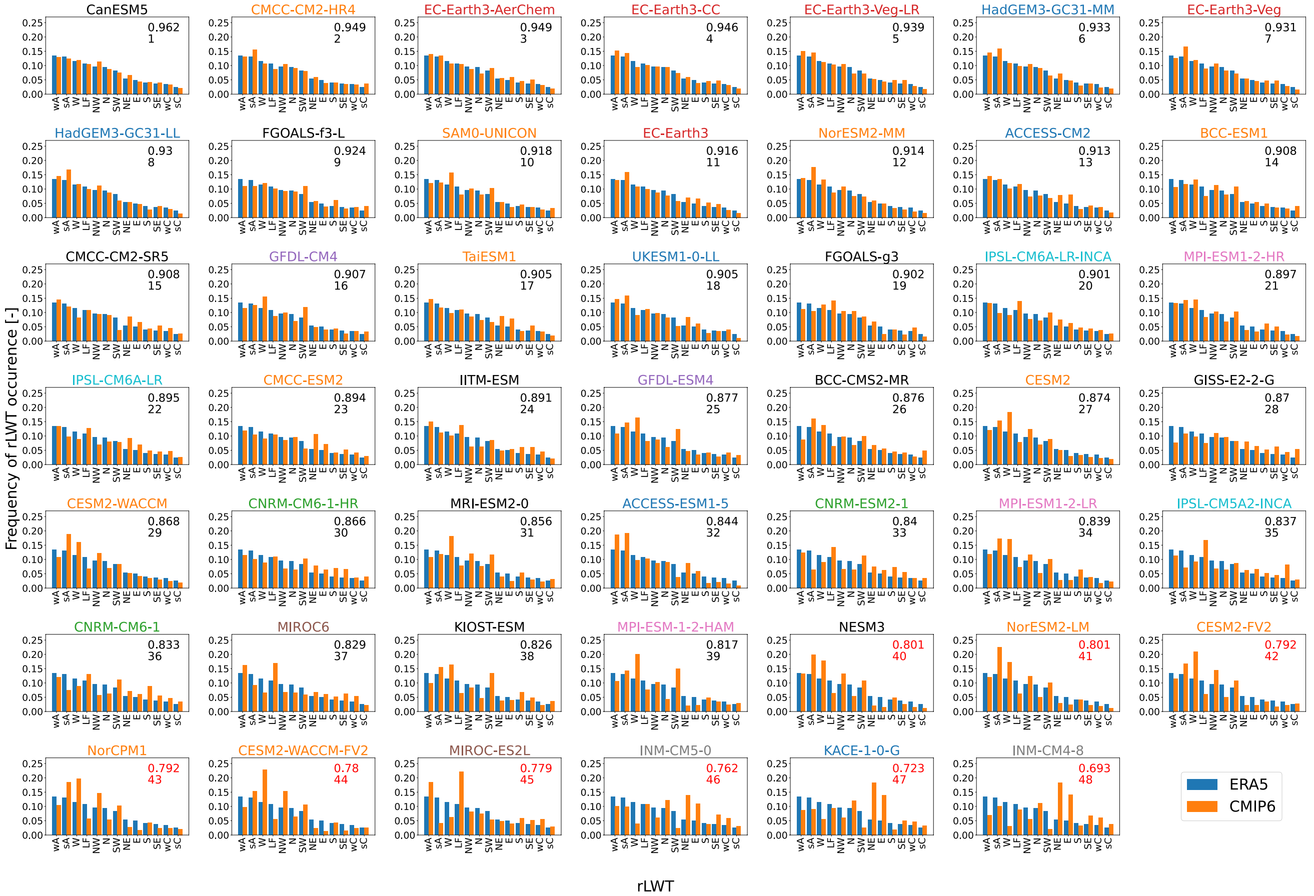


Fig. 15 Extensive overview representing the frequency of the different Lamb Weather Types (LWT) for different CMIP6 models. The Perkins Skill Score as well as its rank compared to the other models is shown. Red values indicate a performance worse than $\mu - \sigma$. The colours of the model names refer to the different model families as given in Table 1

Table 7 Overview of the CMIP6 models used in this work, including the atmosphere components, the nominal resolution in longitude and latitude, and the number of vertical model levels

CMIP6 model	Run	Atmospheric model	Model family	Version	References
ACCESS-CM2	r1i1p1f1	MetUM-HadGEM3-GA7.1 (N96, 192 × 144, 85 lv)	HadGAM/UM	v20210317	Bi et al. (2020)
ACCESS-ESM1-5	r1i1p1f1	HadGAM2 (r1.1, N96, 192 × 145, 38 lv)	HadGAM/UM	v20210318	Ziehn et al. (2020)
BCC-CSM2-MR	r2i1p1f1	BCC-AGCM3-MR (320 × 160, 46 lv)	CAM	v20181216	Wu et al. (2019)
BCC-ESM1	r1i1p1f1	BCC-AGCM3-Chem	CAM	v20181220	Wu et al. (2020)
CanESM5	r1i1p1f1	CanAM5 (T63, T63 Linear Gaussian Grid, 128 × 64, 49 lv)	CanAM	v20190429	Swart et al. (2019)
CESM2	r4i1p1f1	CAM6 (1deg, 288 × 192, 32 lv)	CAM	v20190308	Danabasoglu et al. (2020)
CESM2-FV2	r1i1p1f1	CAM6 (2deg, 144 × 96, 32 lv)	CAM	v20191120	Danabasoglu et al. (2020)
CESM2-WACCM	r1i1p1f1	WACCM6 (1deg, 288 × 192, 70 lv)	CAM	v20190227	Danabasoglu et al. (2020)
CESM2-WACCM-FV2	r1i1p1f1	WACCM6 (2deg, 144 × 96, 70 lv)	CAM	v20191120	Danabasoglu et al. (2020)
CMCC-CM2-HR4	r1i1p1f1	CAM4 (1 deg, 288 × 192, 26 lv)	CAM	v20200904	Cherchi et al. (2019)
CMCC-CM2-SR5	r1i1p1f1	CAM5.3 (1 deg, 288 × 192, 30 lv)	CAM	v20200616	Cherchi et al. (2019)
CMCC-ESM2	r1i1p1f1	CAM5.3 (1deg, 288 × 192, 30 lv)	*CAM*	v20210202	Lovato et al. (2022)
CNRM-CM6-1	r1i1p1f2	Arpege 6.3 (T127, rg with 24572 gp, 91 lv)	ARPEGE	v20190219	Voldoire et al. (2019)
CNRM-CM6-1-HR	r1i1p1f2	Arpege 6.3 (T359, rg with 181724 gp, 91 lv)	ARPEGE	v20191202	Voldoire et al. (2019)
CNRM-ESM2-1	r1i1p1f2	Arpege 6.3 (T127, rg with 24572 gp, 91 lv)	ARPEGE	v20191021	Séférian et al. (2019)
EC-Earth3	r1i1p1f1	IFS cy36r4 (TL255, linearly rg, 512 × 256, 91 lv)	IFS	v20200310	Döscher et al. (2022)
EC-Earth3-AerChem	r1i1p1f1	IFS cy36r4 (TL255, linearly rg, 512 × 256, 91 lv)	IFS	v20200624	Döscher et al. (2022)
EC-Earth3-CC	r1i1p1f1	IFS cy36r4 (TL255, linearly rg, 512 × 256, 91 lv)	IFS	v20210113	Döscher et al. (2022)
EC-Earth3-Veg	r1i1p1f1	IFS cy36r4 (TL255, linearly rg, 512 × 256, 91 lv)	IFS	v20221112	Döscher et al. (2022)
EC-Earth3-Veg-LR	r1i1p1f1	IFS cy36r4 (TL159, linearly rg, 320 × 160, 62 lv)	IFS	v20201123	Döscher et al. (2022)
FGOALS-f3-L	r1i1p1f1	FAMIL2.2 (c96, 360 × 180, 32 lv)	*GAMIL*	v20191019	He et al. (2020)
FGOALS-g3	r1i1p1f1	GAMIL3 (180 × 80, 26 lv)	*GAMIL*	v20190820	Li et al. (2020b)
GFDL-CM4	r1i1p1f1	GFDL-AM4.0.1 (cubed sphere, c96, 360 × 180, 33 lv)	GFDL-AM	v20180701	Held et al. (2019)
GFDL-ESM4	r1i1p1f1	GFDL-AM4.1 (cubed sphere, c96, 360 × 180, 46 lv)	GFDL-AM	v20180701	Dunne et al. (2020)
GISS-E2-2-G	r1i1p1f1	GISS-E2.2 (144 × 90, 102 lv)	GISS-E2	v20191120	Rind et al. (2020)
HadGEM3-GC31-LL	r1i1p1f3	MetUM-HadGEM3-GA7.1 (N96, 192 × 144, 85 lv)	HadGAM/UM	v20191207	Roberts et al. (2019)
HadGEM3-GC31-MM	r1i1p1f3	MetUM-HadGEM3-GA7.1 (N126, 432 × 324, 85 lv)	HadGAM/UM	v20201114	Roberts et al. (2019)
IITM-ESM	r1i1p1f1	IITM-GFSv1 (T62, Linearly rg; 192 × 94 64 lv)	*GFS*	v20201112	Swapna et al. (2015)
INM-CM4-8	r1i1p1f1	INM-AM5-0 (2 × 1,5, 180 × 120, 21 lv)	INM-AM	v20190530	Volodin et al. (2018)
INM-CM5-0	r1i1p1f1	INM-AM5-0 (2 × 1,5, 180 × 120, 73 lv)	INM-AM	v20190610	Volodin et al. (2017)
IPSL-CM5A2-INCA	r1i1p1f1	LMDZ (APv5, 96 × 96, 39 lv)	LMDZ	v20200729	
IPSL-CM6A-LR	r1i1p1f1	LMDZ (NPv6, N96, 144 × 143, 79 lv)	LMDZ	v20190614	Boucher et al. (2020)
IPSL-CM6A-LR-INCA	r1i1p1f1	LMDZ (NPv6, 144 × 143, 79 lv)	LMDZ	v20210216	
KACE-1-0-G	r2i1p1f1	MetUM-HadGEM3-GA7.1 (N96, 192 × 144, 85 lv)	HadGAM/UM	v20200102	Lee et al. (2020a)
KIOST-ESM	r1i1p1f1	GFDL-AM2.0 (cubed sphere, c48, 192 × 96, 32 lv)	GFDL-AM	v20210601	Pak et al. (2021)

Table 7 (continued)

CMIP6 model	Run	Atmospheric model	Model family	Version	References
MIROC6	r1i1p1f1	CCSR AGCM (T85, 256 × 128, 81 lv)	MIROC-AGCM/CCSR	v20191016	Tatebe et al. (2019)
MIROC-ES2L	r1i1p1f2	CCSR AGCM (T42, 128 × 64, 40 lv)	MIROC-AGCM/CCSR	v20200318	Hajima et al. (2020)
MPI-ESM1-2-HR	r1i1p1f1	ECHAM6.3 (spectral T63, 384 × 192, 95 lv)	ECHAM	v20190710	Mauritsen et al. (2019)
MPI-ESM1-2-LR	r1i1p1f1	ECHAM6.3 (spectral T63, 192 × 96, 47 lv)	ECHAM	v20190710	Mauritsen et al. (2019)
MPI-ESM-1-2-HAM	r1i1p1f1	ECHAM6.3 (spectral T63, 192 × 96, 47 lv)	ECHAM	v20190627	Müller et al. (2018)
MRI-ESM2-0	r1i1p1f1	MRI-AGCM3.5 (TL159, 302 × 160, 80 lv)	GSMUV/MRI-AGCM	v20191108	Yukimoto et al. (2019)
NESM3	r1i1p1f1	ECHAM v6.3 (T63, 192 × 96, 47 lv)	ECHAM	v20190812	Cao et al. (2018)
NorCPM1	r1i1p1f1	CAM-OSLO4.1 (2deg, 144 × 96, 26lv)	CAM	v20200724	Bethke et al. (2021)
NorESM2-LM	r1i1p1f1	CAM-OSLO (2deg, 144 × 96, 32 lv)	CAM	v20190815	Seland et al. (2020)
NorESM2-MM	r1i1p1f1	CAM-OSLO (1deg, 288 × 192, 32 lv)	CAM	v20191108	Seland et al. (2020)
SAM0-UNICON	r1i1p1f1	CAM5.3 (288 × 192, 30 lv)	CAM	v20190323	Park et al. (2019)
TaiESM1	r1i1p1f1	TaiAM1 (.9 × 1.25 deg, 288 × 192, 30 lv)	CAM	v20210517	Lee et al. (2020b)
UKESM1-0-LL	r1i1p1f2	MetUM-HadGEM3-GA7.1 (N96, 192 × 144, 85 lv)	HadGAM/UM	v20190715	Sellar et al. (2019)

The downloaded version as well as the reference articles are also included. 'rg' refers to a Reduced Gaussian grid configuration, and 'gp' refers to grid points. 'Model family' follows the grouping made by Brands (2022b) and Brunner et al. (2020), with italics referring to GCMs that did not fulfil the grouping criteria

Appendix 6: Extensive CMIP6 evaluation

The evaluation of the different CMIP6 models was carried out for multiple models. As not all models had all the required data for the heat analysis, they were not included in the paper. However, the complete overview of the evaluation can be found in Fig. 15. Here, the threshold for model exclusion equals 0.844. A detailed description of the atmospheric components of the different CMIP6 models can be found in Table 7. Additionally, the version of the dataset for the historical pressure is given.

Acknowledgements The authors acknowledge Copernicus for the ERA5 reanalysis data. We also acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate model groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF.

Author contributions Fien Serras contributed to the conceptualisation and methodology, the curating of data, the analysis of the data, the visualisation and the writing of the paper. Kobe Vandelanotte, Ruben Borgers and Matthias Demuzere contributed to the methodology and the writing (review and editing) of the paper. Piet Termonia contributed to the funding acquisition and the writing (review and editing) of the paper. Bert Van Schaeybroeck contributed to the methodology, the supervision and the writing (review and editing) of the paper. Nicole van Lipzig contributed to the conceptualisation, the methodology, the data interpretation, the overall supervision of the research project, and the writing (review and editing) of the paper.

Funding This research was funded by BELSPO (grant B2/223/P1/CORDEX.be II). M. Demuzere is supported by European Union's HORIZON Research and Innovation Actions under grant agreement No. 101137851, project CARMINE (Climate-Resilient Development Pathways in Metropolitan Regions of Europe, https://www.carmine-project.eu/).

Availability of data and materials This ERA5 reanalysis dataset can be accessed through the Copernicus Climate Change Service (*C3S*) at ECMWF. The CMIP6 model data can be accessed through the ESGF. The data to recreate the figures of this paper can be found in the KU Leuven Research Data Repository (RDR) at https://doi.org/10.48804/MINSTY.

Code availability The KU Leuven Research Data Repository (RDR) contains the code for the Lamb Weather Type classification at https://doi.org/10.48804/7HHSVV and the scripts to perform the optimization at https://doi.org/10.48804/VINEVJ.

Declarations

Conflict of interest All authors declare that they have no Conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Basara JB, Basara HG, Illston BG et al (2010) The impact of the urban heat island during an intense heat wave in Oklahoma City. Adv Meteorol 2010:1–10. https://doi.org/10.1155/2010/230365

Basu R (2002) Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. Epidemiol Rev 24(2):190–202. https://doi.org/10.1093/epirev/mxf007

Beck C, Philipp A (2010) Evaluation and comparison of circulation type classifications for the European domain. Phys Chem Earth PT A/B/C 35(9):374–387. https://doi.org/10.1016/j.pce.2010.01.001

Bethke I, Wang Y, Counillon F et al (2021) NorCPM1 and its contribution to CMIP6 DCPP. Geosci Model Dev 14(11):7073–7116. https://doi.org/10.5194/gmd-14-7073-2021

Bi D, Dix M, Marsland S et al (2020) Configuration and spin-up of ACCESS-CM2, the new generation Australian community climate and earth system simulator coupled model. J South Hemisph Earth Syst Sci 70(1):225–251. https://doi.org/10.1071/ES19040

Boucher O, Servonnat J, Albright AL et al (2020) Presentation and evaluation of the IPSL-CM6A-LR climate model. J Adv Model Earth Syst 12(7):e2019MS002010. https://doi.org/10.1029/2019MS002010

Brands S (2022a) A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes. Geosci Model Dev 15(4):1375–1411. https://doi.org/10.5194/gmd-15-1375-2022

Brands S (2022b) Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble. Geophys Res Lett 49(23):e2022GL101446. https://doi.org/10.1029/2022GL101446

Brands S, Herrera S, Gutiérrez J (2014) Is Eurasian snow cover in October a reliable statistical predictor for the wintertime climate on the Iberian Peninsula? Int J Climatol 34(5):1615–1627. https://doi.org/10.1002/joc.3788

Brisson E, Demuzere M, Kwakernaak B et al (2011) Relations between atmospheric circulation and precipitation in Belgium. Meteorol Atmos Phys 111(1):27–39. https://doi.org/10.1007/s00703-010-0103-y

Broderick C, Fealy R (2015) An analysis of the synoptic and climatological applicability of circulation type classifications for Ireland. Int J Climatol 35(4):481–505. https://doi.org/10.1002/joc.3996

Brouwers J, Peeters B, Van Steertegem M et al (2015) MIRA klimaatrapport 2015, over waargenomen en toekomstige klimaatveranderinging. Vlaamse Milieumaatschappij i.s.m. KU Leuven , VITO en KMI, Aalst, Belgium

Brunner L, Hegerl GC, Steiner AK (2017) Connecting atmospheric blocking to European temperature extremes in spring. J Clim 30(2):585–594. https://doi.org/10.1175/JCLI-D-16-0518.1

Brunner L, Pendergrass AG, Lehner F et al (2020) Reduced global warming from CMIP6 projections when weighting models by performance and independence. Earth Syst Dyn 11(4):995–1012 https://doi.org/10.5194/esd-11-995-2020. https://esd.copernicus.org/articles/11/995/2020/

Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. Commun Stat Theory Methods 3(1):1–27. https://doi.org/10.1080/03610927408827101

Cao J, Wang B, Yang YM et al (2018) The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation. Geosci Model Dev 11(7):2975–2993. https://doi.org/10.5194/gmd-11-2975-2018

Cherchi A, Fogli PG, Lovato T et al (2019) Global mean climate and main patterns of variability in the CMCC–CM2 coupled model. J Adv Model Earth Syst 11(1):185–209. https://doi.org/10.1029/2018MS001369

Danabasoglu G, Lamarque J, Bacmeister J et al (2020) The community earth system model version 2 (CESM2). J Adv Model Earth Syst 12(2):e2019MS001916. https://doi.org/10.1029/2019MS001916

De B, Tselioudis G, Polvani LM (2022) Improved representation of atmospheric dynamics in CMIP6 models removes climate sensitivity dependence on Hadley cell climatological extent. Atmos Sci Lett 23(3):e1073. https://doi.org/10.1002/asl.1073

Demuzere M, Werner M, Van Lipzig NPM et al (2009) An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns. Int J Climatol 29(12):1796–1810. https://doi.org/10.1002/joc.1821

Di Virgilio G, Ji F, Tam E et al (2022) Selecting CMIP6 GCMs for CORDEX dynamical downscaling: model performance, independence, and climate change signals. Earth's Future 10(4):e2021EF002625. https://doi.org/10.1029/2021EF002625

Dufresne JL, Foujols MA, Denvil S et al (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Clim Dyn 40(9):2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Dunne JP, Horowitz LW, Adcroft AJ et al (2020) The GFDL earth system model version 4.1 (GFDL-ESM 4.1): overall coupled model description and simulation characteristics. J Adv Model Earth Syst 12(11):e2019MS002015. https://doi.org/10.1029/2019MS002015

Döscher R, Acosta M, Alessandri A et al (2022) The EC-Earth3 earth system model for the coupled model intercomparison project 6. Geosci Model Dev 15(7):2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

ECMWF (2016) IFS documentation CY41R2—part IV: physical processes. 4, ECMWF. https://doi.org/10.21957/tr5rv27xu. https://www.ecmwf.int/node/16648

ECMWF (2020) ERA5: 2 metre temperature. https://confluence.ecmwf.int/749 display/CKB/ERA5%3A+2+metre+temperature [Retrieved 2023-11-08]

Epstein Y, Moran DS (2006) Thermal comfort and the heat stress indices. Ind Health 44(3):388–398. https://doi.org/10.2486/indhealth.44.388

ES-DOC (2016) Dataset errata. https://errata.ipsl.fr/static/index.html. Last accessed on 31-Jan-2024

Eyring V, Bony S, Meehl GA et al (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 9(5):1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fan X, Miao C, Duan Q et al (2020) The performance of CMIP6 versus CMIP5 in simulating temperature extremes over the global land surface. J Geophys Res Atmos 125(18):e2020JD033031. https://doi.org/10.1029/2020JD033031

Fernández-Granja JA, Brands S, Bedia J et al (2023) Exploring the limits of the Jenkinson–Collison weather types classification scheme:

a global assessment based on various reanalyses. Clim Dyn 61(3):1829–1845. https://doi.org/10.1007/s00382-022-06658-7

Fischer EM, Oleson KW, Lawrence DM (2012) Contrasting urban and rural heat stress responses to climate change. Geophys Res Lett 39(3):2011GL050576. https://doi.org/10.1029/2011GL050576

Grimalt M, Tomàs M, Alomar G et al (2013) Determination of the Jenkinson and Collison's weather types for the western Mediterranean basin over the 1948–2009 period. Temporal analysis. Atmósfera 26(1):75–94. https://doi.org/10.1016/S0187-6236(13)71063-4

Gualdi S, Somot S, Li L et al (2013) The CIRCE simulations: regional climate change projections with realistic representation of the Mediterranean sea. Bull Am Meteorol Soc 94(1):65–81. https://doi.org/10.1175/BAMS-D-11-00136.1

Hajima T, Watanabe M, Yamamoto A et al (2020) Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. Geosci Model Dev 13(5):2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

He B, Yu Y, Bao Q et al (2020) CAS FGOALS-f3-L model dataset descriptions for CMIP6 DECK experiments. Atm Ocean Sci Lett 13(6):582–588. https://doi.org/10.1080/16742834.2020.1778419

Held IM, Guo H, Adcroft A et al (2019) Structure and Performance of GFDL's CM4.0 Climate Model. J Adv Model Earth Syst 11(11):3691–3727. https://doi.org/10.1029/2019MS001829

Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/qj.3803

Hersbach H, Bell B, Berrisford P et al (2023) ERA5 hourly data on single levels from 1940 to present. https://doi.org/10.24381/cds.adbb2d47. Retrieved 27-Oct-2023

Hoogeveen J, Hoogeveen H (2023) Winds are changing: an explanation for the warming of the Netherlands. Int J Climatol 43(1):354–371. https://doi.org/10.1002/joc.7763

Huth R, Beck C, Philipp A et al (2008) Classifications of atmospheric circulation patterns: recent advances and applications. Ann N Y Acad Sci 1146(1):105–152. https://doi.org/10.1196/annals.1446.019

IPCC (2023) Climate Change 2021—The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1st edn. Cambridge University Press. https://doi.org/10.1017/9781009157896. https://www.cambridge.org/core/product/identifier/9781009157896/type/book

Jenkinson AF, Collison FP (1977) An initial climatology of gales over the north sea. Synopt Climatol Branch Memorandum 62:18

Jones PD, Hulme M, Briffa KR (1993) A comparison of Lamb circulation types with an objective classification scheme. Int J Climatol 13(6):655–663. https://doi.org/10.1002/joc.3370130606

Jones PD, Harpham C, Briffa KR (2013) Lamb weather types derived from reanalysis products. Int J Climatol 33(5):1129–1139. https://doi.org/10.1002/joc.3498

Kim YH, Min SK, Zhang X et al (2020) Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. Weather Clim Extrem 29:100269. https://doi.org/10.1016/j.wace.2020.100269

Knutti R, Furrer R, Tebaldi C et al (2010) Challenges in combining projections from multiple climate models. J Clim 23(10):2739–2758. https://doi.org/10.1175/2009JCLI3361.1

Kotlarski S, Keuler K, Christensen OB et al (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7(4):1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Laaidi K, Zeghnoun A, Dousset B et al (2012) The impact of heat Islands on mortality in Paris during the August 2003 heat wave. Environ Health Perspect 120(2):254–259. https://doi.org/10.1289/ehp.1103532

Lamb HH (1950) Types and spells of weather around the year in the British Isles: annual trends, seasonal structure of the year, singularities. Q J R Meteorol Soc 76(330):393–429. https://doi.org/10.1002/qj.49707633005

Lamb HH (1972) British isles weather types and a register of the daily sequence of circulation patterns 1861–1971. Geophys Mem 1972:1

Lee J, Kim J, Sun MA et al (2020) Evaluation of the Korea meteorological administration advanced community earth-system model (K-ACE). Asia Pac J Atmos Sci 56(3):381–395. https://doi.org/10.1007/s13143-019-00144-7

Lee WL, Wang YC, Shiu CJ et al (2020) Taiwan earth system model version 1: description and evaluation of mean state. Geosci Model Dev 13(9):3887–3904. https://doi.org/10.5194/gmd-13-3887-2020

Li D, Yuan J, Kopp RE (2020) Escalating global exposure to compound heat-humidity extremes with warming. Environ Res Lett 15(6):064003. https://doi.org/10.1088/1748-9326/ab7d04

Li H, Sheffield J, Wood EF (2010) Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. J Geophys Res 115:2009JD012882. https://doi.org/10.1029/2009JD012882

Li L, Yu Y, Tang Y et al (2020) The flexible global ocean–atmosphere–land system model grid-point version 3 (FGOALS-g3): description and evaluation. J Adv Model Earth Syst 12(9):e2019MS002012. https://doi.org/10.1029/2019MS002012

Liang X, Kunkel KE, Meehl GA et al (2008) Regional climate models downscaling analysis of general circulation models present climate biases propagation into future change projections. Geophys Res Lett 35(8):2007GL032849. https://doi.org/10.1029/2007GL032849

Lovato T, Peano D, Butenschön M et al (2022) CMIP6 simulations with the CMCC earth system model (CMCC-ESM2). J Adv Model Earth Syst 14(3):e2021MS002814. https://doi.org/10.1029/2021MS002814

Masterton JM, Richardson F (1979) Humidex: a method of quantifying human discomfort due to excessive heat and humidity. Environment Canada, Atmospheric Environment

Mauritsen T, Bader J, Becker T et al (2019) Developments in the MPI-M earth system model version 12 (MPI-ESM12) and its response to increasing CO_2. J Adv Model Earth Syst 11(4):998–1038. https://doi.org/10.1029/2018MS001400

McSweeney CF, Jones RG, Lee RW et al (2015) Selecting CMIP5 GCMs for downscaling over multiple regions. Clim Dyn 44(11):3237–3260. https://doi.org/10.1007/s00382-014-2418-8

Morice C, Kennedy J, Rayner N et al (2021) An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. J Geophys Res (Atmos). https://doi.org/10.1029/2019JD032361. Accessed 2023-07-14

Müller WA, Jungclaus JH, Mauritsen T et al (2018) A higher-resolution version of the max Planck institute earth system model (MPI-ESM1.2-HR). J Adv Model Earth Syst 10(7):1383–1413. https://doi.org/10.1029/2017MS001217

Otero N, Sillmann J, Butler T (2018) Assessment of an extended version of the Jenkinson–Collison classification on CMIP5 models over Europe. Clim Dyn 50(5–6):1559–1579. https://doi.org/10.1007/s00382-017-3705-y

Pak G, Noh Y, Lee MI et al (2021) Korea institute of ocean science and technology earth system model and its simulation characteristics. Ocean Sci J 56(1):18–45. https://doi.org/10.1007/s12601-021-00001-7

Park S, Shin J, Kim S et al (2019) Global climate simulated by the Seoul national university atmosphere model version 0 with a unified convection scheme (SAM0-UNICON). J Clim 32(10):2917–2949. https://doi.org/10.1175/JCLI-D-18-0796.1

Perkins SE, Pitman AJ, Holbrook NJ et al (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J Clim 20(17):4356–4376. https://doi.org/10.1175/JCLI4253.1

Powell MJD (1994) A direct search optimization method that models the objective and constraint functions by linear interpolation. In: Gomez S, Hennart JP (eds) Advances in optimization and numerical analysis. Springer, London, pp 51–67. https://doi.org/10.1007/978-94-015-8330-5_4

Rind D, Orbe C, Jonas J et al (2020) GISS model E2.2: a climate model optimized for the middle atmosphere-model structure, climatology, variability, and climate sensitivity. J Geophys Res Atmos 125(10):e2019JD032204. https://doi.org/10.1029/2019JD032204

Roberts MJ, Baker A, Blockley EW et al (2019) Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. Geosci Model Dev 12(12):4999–5028. https://doi.org/10.5194/gmd-12-4999-2019

Rohde RA, Hausfather Z (2020) The Berkeley earth land/ocean temperature record. Sci Data Earth Syst. https://doi.org/10.5194/essd-12-3469-2020. [Accessed 2023-11-13]

Sanderson MG, Hemming DL, Betts RA (2011) Regional temperature and precipitation changes under high-end (≥4° C) global warming. Philos Trans R Soc A Math Phys Eng Sci 369(1934):85–98. https://doi.org/10.1098/rsta.2010.0283

Seland Ø, Bentsen M, Olivié D et al (2020) Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. Geosci Model Dev 13(12):6165–6200. https://doi.org/10.5194/gmd-13-6165-2020

Sellar AA, Jones CG, Mulcahy JP et al (2019) UKESM1: description and evaluation of the UK Earth system model. J Adv Model Earth Syst 11(12):4513–4558. https://doi.org/10.1029/2019MS001739

Seneviratne SI, Donat MG, Pitman AJ et al (2016) Allowable CO_2 emissions based on regional and impact-related climate targets. Nature 529(7587):477–483. https://doi.org/10.1038/nature16542

Sobolowski S, Somot S, Fernandez J et al (2023) EURO-CORDEX CMIP6 GCM selection and ensemble design: best practices and recommendations. Zenodo. https://doi.org/10.5281/zenodo.7673400

Swapna P, Roxy MK, Aparna K et al (2015) The IITM earth system model: transformation of a seasonal prediction model to a long-term climate model. Bull Am Meteorol Soc 96(8):1351–1367. https://doi.org/10.1175/BAMS-D-13-00276.1

Swart NC, Cole JNS, Kharin VV et al (2019) The Canadian earth system model version 5 (CanESM5.0.3). Geosci Model Dev 12(11):4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Séférian R, Nabat P, Michou M et al (2019) Evaluation of CNRM earth system model, CNRM-ESM2-1: role of earth system processes in present-day and future climate. J Adv Model Earth Syst 11(12):4182–4227. https://doi.org/10.1029/2019MS001791

Tatebe H, Ogura T, Nitta T et al (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci Model Dev 12(7):2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Tebaldi C, Knutti R (2018) Evaluating the accuracy of climate change pattern emulation for low warming targets. Environ Res Lett 13(5):055006. https://doi.org/10.1088/1748-9326/aabef2

Tomczyk AM, Owczarek M (2020) Occurrence of strong and very strong heat stress in Poland and its circulation conditions. Theor Appl Climatol 139(3–4):893–905. https://doi.org/10.1007/s00704-019-02998-3

Trigo RM, DaCamara CC (2000) Circulation weather types and their influence on the precipitation regime in Portugal. Int J Climatol 20(13):1559–1581. https://doi.org/10.1002/1097-0088(20001115)20:13<1559::AID-JOC555>3.0.CO;2-5

Van De Walle J, Brousse O, Arnalsteen L et al (2022) Lack of vegetation exacerbates exposure to dangerous heat in dense settlements in a tropical African city. Environ Res Lett 17(2):024004. https://doi.org/10.1088/1748-9326/ac47c3

Vautard R, Gobiet A, Sobolowski S et al (2014) The European climate under a 2 °C global warming. Environ Res Lett 9(3):034006. https://doi.org/10.1088/1748-9326/9/3/034006

Vautard R, Cattiaux J, Happé T et al (2023) Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. Nat Commun 14(1):6803. https://doi.org/10.1038/s41467-023-42143-3

Voldoire A, Saint-Martin D, Sénési S et al (2019) Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. J Adv Model Earth Syst 11(7):2177–2213. https://doi.org/10.1029/2019MS001683

Volodin EM, Mortikov EV, Kostrykin SV et al (2017) Simulation of the present-day climate with the climate model INMCM5. Clim Dyn 49(11–12):3715–3734. https://doi.org/10.1007/s00382-017-3539-7

Volodin EM, Mortikov EV, Kostrykin SV et al (2018) Simulation of the modern climate using the INM-CM48 climate model. Russ J Numer Anal Math Model 33(6):367–374. https://doi.org/10.1515/rnam-2018-0032

World Meteorological Organization (2017) WMO guidelines on the calculation of climate normals. World Meteorological Organization, Geneva

Wouters H, De Ridder K, Poelmans L et al (2017) Heat stress increase under climate change twice as large in cities as in rural areas: a study for a densely populated midlatitude maritime region. Geophys Res Lett 44(17):8997–9007. https://doi.org/10.1002/2017GL074889

Wu T, Lu Y, Fang Y et al (2019) The Beijing climate center climate system model (BCC-CSM): the main progress from CMIP5 to CMIP6. Geosci Model Dev 12(4):1573–1600. https://doi.org/10.5194/gmd-12-1573-2019

Wu T, Zhang F, Zhang J et al (2020) Beijing climate center earth system model version 1 (BCC-ESM1): model description and evaluation of aerosol simulations. Geosci Model Dev 13(3):977–1005. https://doi.org/10.5194/gmd-13-977-2020

Yukimoto S, Kawai H, Koshiro T et al (2019) The meteorological research institute earth system model version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component. J Meteorol Soc Jpn Ser II 97(5):931–965. https://doi.org/10.2151/jmsj.2019-051

Zelinka MD, Myers TA, McCoy DT et al (2020) Causes of higher climate sensitivity in CMIP6 models. Geophys Res Lett 47(1):e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhang HM, Huang B, Lawrimore J et al (2023) NOAA global surface temperature dataset (NOAAGlobalTemp), version 5.1, NOAA National Centers for Environmental Information. https://doi.org/10.25921/9qth-2p70. Accessed 2023-07-14

Ziehn T, Chamberlain MA, Law RM et al (2020) The Australian earth system model: ACCESS-ESM1.5. J South Hemisph Earth Syst Sci 70(1):193–214. https://doi.org/10.1071/ES19035