Quarterly Journal of the
Royal Meteorological Society | RMetS

Forced trends and internal variability in climate change projections of extreme European windstorm frequency and severity

Matthew D. K. Priestley[1] | David B. Stephenson[1] | Adam A. Scaife[1,2] | Daniel Bannister[3] | Christopher J. T. Allen[4] | David Wilkie[4]

[1]Department of Mathematics and Statistics, University of Exeter, Exeter, UK

[2]Met Office, Exeter, UK

[3]WTW Research Network, WTW, London, UK

[4]Model Research and Evaluation, Gallagher Re, London, UK

Correspondence
Matthew D. K. Priestley, Department of Mathematics and Statistics, University of Exeter, Exeter, UK.
Email: m.priestley@exeter.ac.uk

Abstract

Climate change projections of European windstorm damages are highly uncertain because of different climate model responses and large internal variability. This study uses generalized linear models and a weighted median estimation to optimally extract forced trends in a number of European windstorm metrics. Footprints of windstorms associated with extratropical cyclones are created for an ensemble of models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) across a full transient time series from 1980 to 2100. Trends are assessed over time, but also as a function of global mean surface temperature changes. Trends in aggregate severity are attributed to changes in storm average severity, frequency, and area impacted, with changes in area being the dominant driver of changes to average storm severity. Confidence in the findings is assessed, with high confidence of declines in frequency for southern and northern Europe, medium confidence of an increase in average windstorm severity for parts of northwestern Europe, and low confidence of any changes for eastern Europe. A 15-member ensemble of the MPI-ESM1-2-LR model is used to assess internal variability. Trends between individual members can vary significantly; however, the uncertainty due to internal variability in the 15-member ensemble is generally only 50% of that in the multimodel ensemble of CMIP6 models for aggregate severity. With largest uncertainty coming from model differences, a large proportion of uncertainty in future windstorms is therefore potentially reducible with modelling advances.

KEYWORDS
Climate change trends, European windstorm, internal variability, model variability, storm severity, weighted median

1 | INTRODUCTION

Extratropical cyclones (ETCs) commonly cause significant impacts across Europe due to the accompanying strong winds (*windstorms*; Ulbrich *et al.*, 2001), and individual events cause damages in excess of €10 billion (Barredo, 2010; Cusack, 2023). Owing to the potentially significant economic and insured impact of these windstorms, it is important to consider how these may evolve in the future under the influence of a warming climate. ETCs tend to show a decrease in frequency in the future (Bengtsson *et al.*, 2006; Geng & Sugi, 2003; Mizuta *et al.*, 2011); however, in current-generation general circulation models (GCMs) most models indicate an extension of the North Atlantic storm track, and therefore increase in frequency for western Europe (Harvey *et al.*, 2020; Priestley & Catto, 2022a; Zappa *et al.*, 2013). These changes are commonly the most pronounced in higher forcing scenarios, with consistent model changes also often only being apparent in these high forcing scenarios, which is likely a result of the high interannual variability present in European storminess (Feser *et al.*, 2015).

Recent studies using the Coupled Model Intercomparison Project Phase 6 (CMIP6) models, such as Severino *et al.* (2024) and Little *et al.* (2023), have found increases in storminess for the central latitudes of Europe, covering countries such as the United Kingdom, France, Germany, and Benelux, with decreases for southern, northern, and eastern Europe. These findings were consistent with other previous single- and multi-model assessments (e.g. Donat *et al.*, 2011; Feser *et al.*, 2015; Leckebusch *et al.*, 2007; Mölter *et al.*, 2016; Pinto *et al.*, 2007).

The main uncertainties in future trend assessments are from model variability, internal variability, and scenario uncertainty (Hawkins & Sutton, 2009). Model variability arises from the different formulation of the equations of motion by different models, and how this affects the representation of physical processes. This can result in considerable variation in future projections of the storm tracks and associated impacts (Little *et al.*, 2023). Internal variability is often the dominant source of uncertainty at short time-scales (Blanusa *et al.*, 2023; Deser *et al.*, 2012; Hawkins & Sutton, 2009), but little attention has been paid to its role on projections of European windstorms. Internal variability has been shown to have a large impact on the representation and trend of the North Atlantic Oscillation (Deser *et al.*, 2017) and tends to arise as a result of the chaotic nature of the atmospheric circulation (Deser *et al.*, 2014; Scaife *et al.*, 2009). As windstorm assessments commonly only use a single ensemble member (e.g. Little *et al.*, 2023), it is therefore important to quantify the impact of internal variability on projections and how this compares with variability from other sources, such as the choice of climate model.

GCMs generally represent ETCs, and their associated impact features well (Catto *et al.*, 2011; Priestley & Catto, 2022b). However, most models partaking in CMIP6 (Eyring *et al.*, 2016) do not have resolution high enough to represent mesoscale features such as sting jets (Clark & Gray, 2018), and may not be able to inform as to how peak loss potential will change. Despite this, insights into changes in European storminess can still be made. Statistical methods that describe the change in the underlying severity distribution with climate change (*forced trends*) can offer a new perspective over previous studies (e.g. Little *et al.*, 2023; Severino *et al.*, 2024).

Despite general consistencies in patterns of change, the magnitude of changes and robustness of findings are often hard to ascertain owing to a number of factors. These factors often arise from varying methodological choices, which are summarized as follows:

- Studies often use single models (Pinto *et al.*, 2007) or single climate scenarios (Leckebusch *et al.*, 2007), and therefore results may be unique to the model of choice or the climate change signal.

- There is often inconsistency on the features being examined. To focus on ETC-driven windstorms, assessments should consider wind features (or footprints) directly associated with these weather systems, which is not the case in Severino *et al.* (2024). Furthermore, there is often sensitivity to the way in which ETCs are defined/identified; however, tracking schemes are often consistent for the most extreme ETCs (Neu *et al.*, 2013).

- Changes are often quantified either for all of Europe or arbitrarily chosen geographic regions (e.g. Little *et al.*, 2023; Zappa *et al.*, 2013), none of which are consistent with one another. For findings to be applicable to (re)insurers with geographically specific portfolios and other socio-economic decision-makers, these assessments should be made for more clearly defined geopolitical boundaries.

- Changes are often assessed in a time-slice approach (e.g. Leckebusch *et al.*, 2007; Little *et al.*, 2023) with the choice of period for the future and the historical baseline often being unique to each study. All models have different representations of interannual, decadal, and multidecadal variability (Woollings *et al.*, 2015). Therefore, an assessment of a full transient time series should be made to account for this.

- Signals of change are dependent on how one decides to combine a multimodel ensemble. There are numerous approaches and challenges to doing this (Knutti *et al.*, 2010), with metrics such as the mean often being heavily influenced by outlier models. All models have

their own individual biases (Palmer *et al.*, 2023), which will impact their projection. Therefore, consideration should be made as to how to combine a model ensemble that is the best representation of the model signal and also accounts for confidence in each model's projection.

The aforementioned points clearly demonstrate the need for a multi-socioeconomic pathways model, multi-scenario assessment of European windstorms driven by ETCs that are quantified for specific geographic regions of Europe. This assessment should examine a full transient climate model experiment and not rely on time-slice analysis. Furthermore, an ensemble should be combined in a way that accounts for model biases and resulting outliers. With a coherent experiment design addressing all these factors, an assessment of confidence in changes of windstorm severity and impact can be made for different regions of Europe.

Taking on board all these open questions, the science questions that this study will address are as follows:

1. Can footprints of European windstorms be created to assess changes in impacts through the 21st century?
2. What consistent trends in windstorm behaviour can be identified across various climate models under different future climate scenarios?
3. To what extent can confidence be attributed to the projections of windstorm behaviour changes based on multimodel ensemble outputs?
4. What is the relative importance of model variability and internal variability in the uncertainty in future trends from climate change?

We set out our data analysis plan and statistical method for assessing the *forced trends* in a number of different European windstorm metrics in Section 2. Our assessment will be made on models from a number of shared socio-economic pathways (O'Neill *et al.*, 2014) and will conclude with an assessment of confidence in changes of the different windstorm metrics for different geographic regions of Europe.

2 | DATA & METHODS

2.1 | CMIP6 data

Analysis is performed on models that are part of the CMIP6 (Eyring *et al.*, 2016) and ScenarioMIP (O'Neill *et al.*, 2016) coupled atmosphere–ocean experiments. In total, 13 different models are used (Table 1), as these are the only models to provide data for the required experiments at the necessary temporal resolutions for cyclone tracking and footprint creation across the historical period and the different future scenarios. Two different future scenarios, or shared socioeconomic pathways (SSPs; O'Neill *et al.*, 2014), are analysed. These are SSP2-45 and SSP5-85, which are the *middle of the road* and *fossil- fuelled development* scenarios respectively and feature differently evolving climate forcing throughout the 21st century—(see O'Neill *et al.*, 2016; Riahi *et al.*, 2017, for full details). In total, 20 realizations are analysed; this is a result of some models providing output for both the SSP2-45 and SSP5-85 scenarios (Table 1). For each model, only a single ensemble member (variant label r1i1p1f1, or lowest available) is used

TABLE 1 Table of models used in this study and indicators of simulations analysed.

| Model name | Modelling centre | Simulations available | | | Ensemble members |
		Historical	SSP2-45	SSP5-85	
BCC-CSM2-MR	Beijing Climate Center	X	X	X	1
CMCC-CM2-SR5	Euro-Mediterranean Center on Climate Change	X		X	1
CMCC-ESM2	Euro-Mediterranean Center on Climate Change	X		X	1
CNRM-CM6-1-HR	Centre National de Recherches Météorologiques	X		X	1
EC-Earth3	EC-Earth-Consortium	X	X	X	1
GISS-E2-1-G	NASA Goddard Institute for Space Studies	X	X	X	1
HadGEM3-GC31-MM	Met Office Hadley Centre	X		X	1
KACE-1-0-G	National Institute of Meteorological Sciences/ Korean Meteorological Administration	X	X		1
MIROC6	JAMSTEC, AORI, NIES, and R-CCS	X	X	X	1
MPI-ESM1-2-HR	Max Planck Institute for Meteorology	X	X	X	1
MPI-ESM1-2-LR	Max Planck Institute for Meteorology	X	X	X	15
MRI-ESM2-0	Meteorological Research Institute, Japan	X	X	X	1
NESM3	Nanjing University of Information, Science, and Technology	X		X	1

to ensure that all models have equal weight. Variables analysed are six-hourly 850 hPa zonal and meridional wind (for identifying and tracking cyclones), daily surface maximum wind speed (for constructing footprints), and monthly averaged surface temperature (for quantify rate of global warming in the different SSP experiments).

The December–February (DJF) period is analysed for just the Northern Hemisphere, representing the winter season. The historical period uses the years 1979–2014, and in each of the future scenarios the years 2015–2100 are available. In addition, a multi-ensemble analysis of one model is performed to quantify the magnitude of internal variability. This is done using the MPI-ESM1-2-LR model, with 15 ensemble members (a similar number to the multi-model set) being used (variant labels r1i1p1f1–r15i1p1f1). This is the only model to provide output of the required variables for more than one ensemble member for the historical and SSP5-85 experiments. The ensemble members are all initialized in 1850 with different initial conditions (Mauritsen *et al.*, 2019). All members receive the same forcing from greenhouse gases and aerosols.

2.2 | European Centre of Medium-range Weather Forecasts Reanalysis v. 5 (ERA5)

The European Centre of Medium-range Weather Forecasts Reanalysis v. 5 (ERA5) (Hersbach *et al.*, 2020) is used in this work as the reference for real-world atmospheric variability and is used to calculate the gust scaling factors to be applied to the CMIP6 model surface daily maximum wind speeds (see later). ERA5 data are available from January 1950 to September 2023; however, the period 1979–2014 is used to begin with the satellite era and have an endpoint consistent with the CMIP6 historical experiments. ERA5 data have horizontal resolution of 0.28° × 0.28° (31 km). Instantaneous data every 6 hr is used for the purposes of cyclone identification and tracking, and daily maximum surface gusts are constructed from hourly output for footprint creation. Gusts in ERA5 are a parametrized variable, which are calculated at every time step as a combination of the instantaneous 10-m wind speed, the turbulent gustiness, and a contribution from convective processes (Bechtold & Bidlot, 2009; Minola *et al.*, 2020; Panofsky *et al.*, 1977). For ERA5 and the CMIP6 models described herein, all model data are regridded to a common resolution of 2°x2° for analysis following footprint creation.

2.3 | Cyclone tracking and footprint creation

To associate windstorms with ETCs, the cyclones themselves first need to be identified. The identification and tracking algorithm TRACK (Hodges, 1994, 1995, 1999) is used and takes six-hourly relative vorticity at 850 hPa as an input. The method first spectrally truncates the vorticity field to T42 resolution and removes the influence of planetary scale waves by masking wave numbers less than 5, which allows for data of differing input resolution to be easily compared. To ensure only mobile, well-developed, and long-lived cyclones are analysed, each track must persist for at least 48 hr and travel at least 1,000 km from its point of origin. Each cyclone must have a maximum
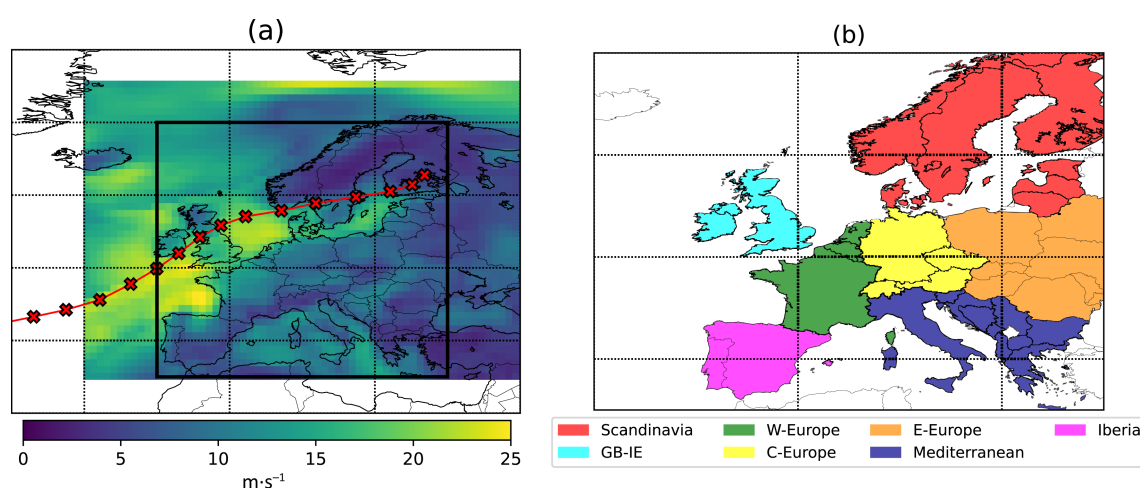
FIGURE 1 Windstorm footprints and regions used in this study. (a) Windstorm footprint of storm 8 in the 1980–1981 season from MPI-ESM1-2-LR. Shading indicates raw daily maximum surface windspeed (sfcWindMax) footprint. Red crosses indicate six-hourly positions of the extratropical cyclone identified. Black square indicates the European region all cyclone tracks must pass through for a footprint to be created. (b) Groupings of countries used for analysis in this study. Countries are grouped into regions following the approach of Severino *et al.* (2024)

vorticity of at least 1×10^{-5} s^{-1}. As the focus of this analysis is for windstorms affecting Europe, a condition is applied to tracks in that they must pass through a region defined by the latitude–longitude bounds of 35°N–70°N, 10°W–30°E (see black box in Figure 1).

Footprints are constructed following the protocol set out in the XWS project (Roberts *et al.*, 2014). The footprint of a windstorm is defined as the maximum 3-s gust at each grid point over a 72-hr period during which the cyclone track passes through the domain. As CMIP6 models do not output 3-s gusts, footprints are created using the daily maximum near-surface (10 m) wind speed (sfcWindMax). The 72-hr period has its central day (day 0) when the cyclone track has its maximum wind speed over land that is within 10° of the identified track point. The resultant footprint uses the two days either side of this day. Footprints for each track are created for all of Europe and the eastern North Atlantic (e.g., Figure 1a).

In subsequent analysis, each footprint is masked to specific geographic areas, which are defined through country groupings. This is due to differences in the cyclone/wind gust climatology and due to the different impacts each region experiences. These groupings are also helpful for applications such as (re)insurance and planning. The

groupings we use follow those proposed by Severino *et al.* (2024). The core regions are shown in Figure 1b and are defined as follows.

- Central Europe (CEU): Austria, Czechia, Germany, Switzerland.
- Eastern Europe (EEU): Belarus, Hungary, Poland, Republic of Moldova, Romania, Slovakia, Ukraine.
- GB/IE: Great Britain, Ireland.
- Iberia (IB): Spain, Portugal.
- Mediterranean (MED): Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Italy, Greece, Macedonia, Malta, Montenegro, San Marino, Serbia, Slovenia.
- Scandinavia (SC): Denmark, Estonia, Finland, Latvia, Lithuania, Norway, Sweden.
- Western Europe (WEU): Belgium, France, Luxembourg, Netherlands.

A region covering all of Europe (EU; all listed countries) is also considered.

The CMIP6 models all have biases in their near-surface wind speeds relative to ERA5 (Carvalho *et al.*, 2021);
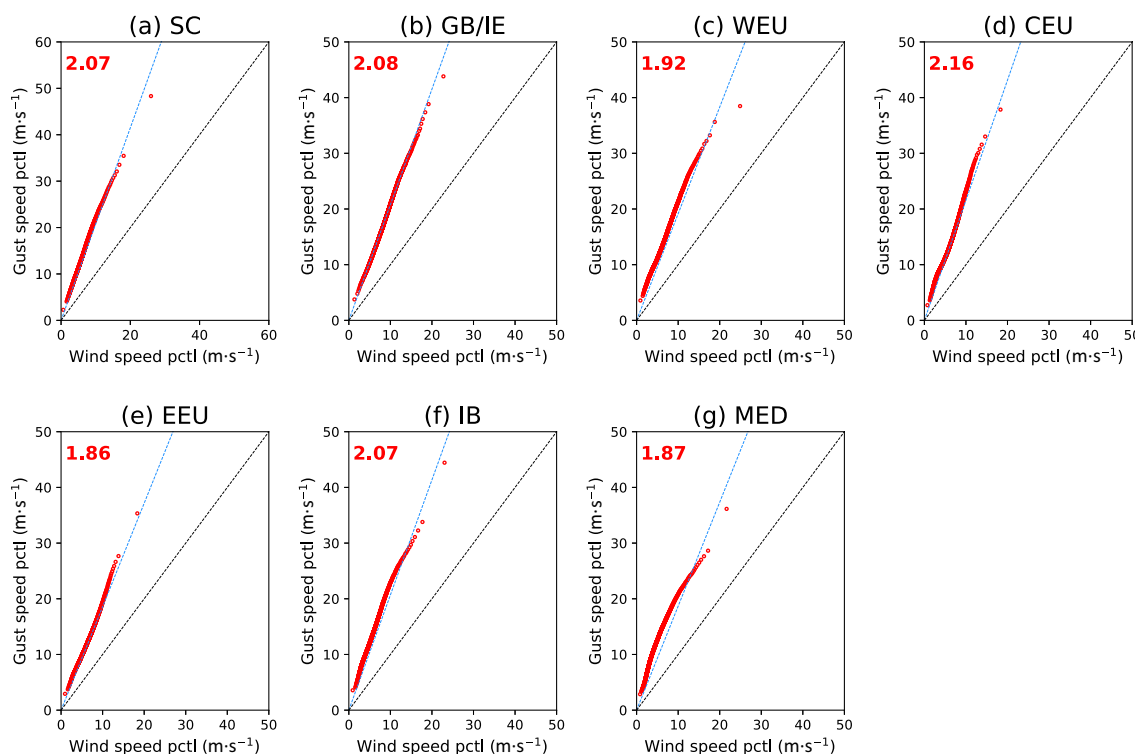
FIGURE 2 Wind-gust speed scaling factors derived from European Centre of Medium-range Weather Forecasts Reanalysis v. 5 (ERA5) for (a) Scandinavia (SC), (b) Great Britain and Ireland (GB/IE), (c) western Europe (WEU), (d) central Europe (CEU), (e) eastern Europe (EEU), (f) Iberia (IB), and (g) Mediterranean (MED). Red dots indicate quantiles of the wind speed and wind gust footprint distributions from 0.5 to 99.5 in steps of 0.5. Units are m·s^{-1}. Dashed blue line is the ratio of the standard deviations of the two distributions, and the black dashed line is the 1:1 line. Inset red text is the slope of the dashed blue line, or the scaling factor from wind speed to wind gust. pctl: percentile.

however, no bias correction is applied in order to take each model's projection at face value. To have footprints for gust speeds instead of daily maximum wind speeds from the CMIP6 models, a scaling factor is applied that has been calculated from ERA5. To match CMIP6, ERA5 daily maximum wind and gust speeds are created from the one-hourly data. Footprints are created for all European cyclones using both daily maximum gusts and wind speeds. By aggregating all grid points within our Europe regions (Figure 1b) and calculating the quantiles of the resultant distributions, a linear scaling factor to convert wind speed to gust speed is obtainable (Figure 2). For all regions the scaling factor is between 1.8 and 2.2 (Figure 2). Taking GB/IE as an example, its scaling factor of 2.08 means that a $10\,m\cdot s^{-1}$ wind speed in a CMIP6 model would scale to a $20.8\,m\cdot s^{-1}$ gust for use in our severity estimations. These scaling values are largely in line with other studies (e.g. Born et al., 2012; Tyner et al., 2015).

2.4 | Storm severity indices

Impacts associated with windstorms are quantified using the storm severity index (SSI) introduced by Klawa and Ulbrich (2003), which has been used extensively in numerous studies (see Karremann et al., 2014; Leckebusch et al., 2007; Little et al., 2023; Priestley et al., 2018). In the Klawa and Ulbrich (2003) formulation of the SSI, a wind gust threshold is applied above which damages begin to occur. The 98th percentile of the local wind (or gust) speed distribution is commonly used as this threshold (Bloomfield et al., 2023; Priestley et al., 2018). As the daily maximum wind speed is used here, anything below the daily maximum at a location is excluded, and therefore the distribution contains very few low wind speeds. Therefore, the 95th percentile of the daily maximum gust speed at each grid point is used as the SSI threshold (Figure 3). For each model we use its own 95th percentile to account for variability in model wind distributions and associated biases. Most models have a 95th percentile that is lower than ERA5 (Figure 3), with some models being more than $10\,m\cdot s^{-1}$ lower (see also Carvalho et al., 2021). In ERA5 this threshold has values of $20–30\,m\cdot s^{-1}$ across most of northwestern Europe.

There are several factors that affect changes in an aggregate SSI (the sum of all SSIs in a winter season). Any changes are a combination of a change in windstorm frequency, windstorm intensity, and also the footprint area. Increases in any of these three factors will result
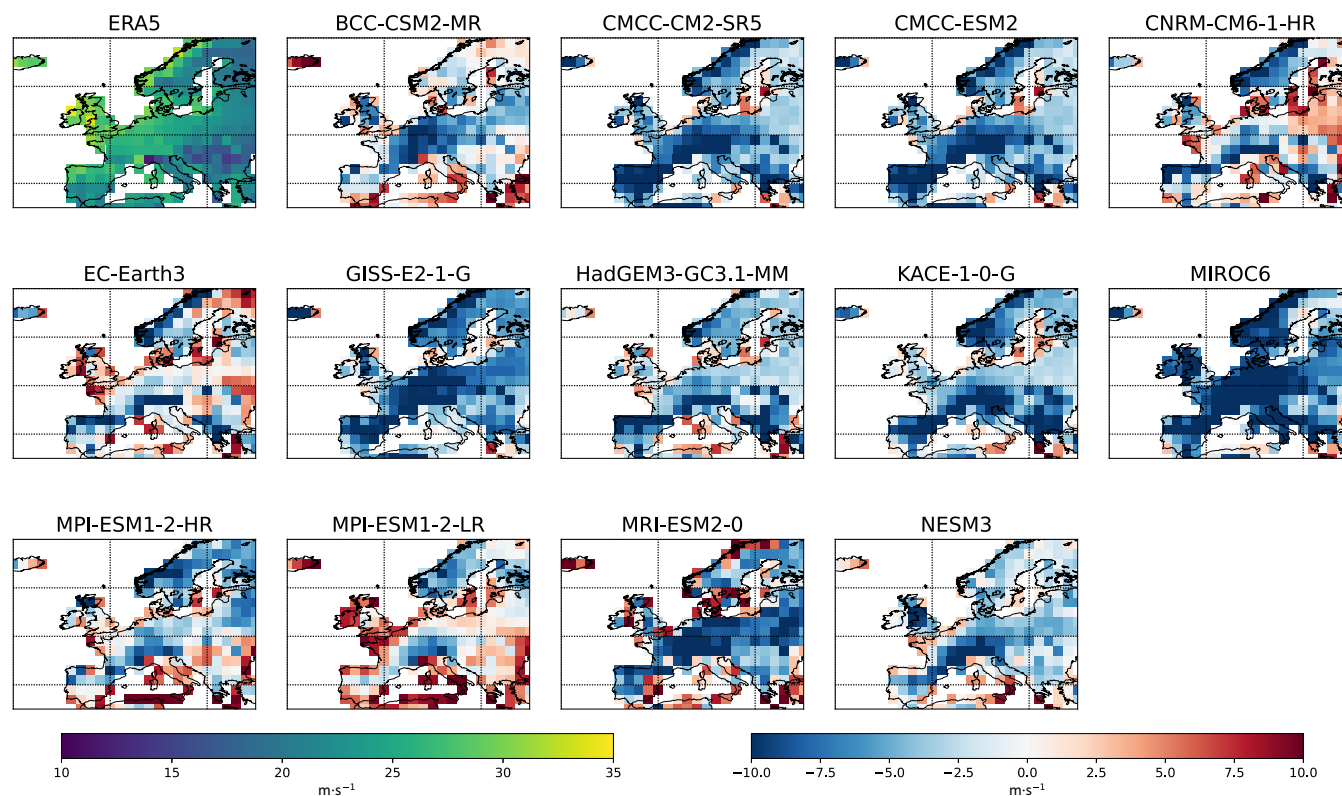
FIGURE 3 The 95th percentile gust threshold used in storm severity index calculations for European Centre of Medium-range Weather Forecasts Reanalysis v. 5 (ERA5) and all Coupled Model Intercomparison Project Phase 6 (CMIP6) models analysed. CMIP6 models are shown as differences relative to ERA5. Units are $m\cdot s^{-1}$.

in an increase in aggregate severity. We examine several variants of the Klawa and Ulbrich (2003) SSI to assess the contribution of each of the aforementioned factors to changes in the aggregate SSI.

At each grid point (x, y) in a defined European region, we can define a gust speed at that location for the ith footprint in any given year as $V_i(x, y)$, and the 95th percentile threshold as $V_{95}(x, y)$. This results in an adapted version of the Klawa and Ulbrich (2003) SSI as

$$L_i = \sum_{V_i(x,y) > V_{95}(x,y)} \left(\frac{V_i(x, y)}{V_{95}(x, y)} - 1 \right)^3. \tag{1}$$

The winter-season aggregate SSI (L) is simply the sum of the L_i for all the footprints from $i = 1, 2, \ldots, n$:

$$L = \sum_{i=1}^{n} L_i. \tag{2}$$

The average SSI per winter ($\overline{L}$) is the mean of all the season footprints (L_i) and is $\overline{L} = L/n$. The area of the footprint (A_i) is defined as the total number of above-threshold grid points of the ith footprint:

$$A_i = \sum_{V_i(x,y) > V_{95}(x,y)} 1, \tag{3}$$

and following the approach in Equation (2) the winter aggregate area (A) is

$$A = \sum_{i=1}^{n} A_i, \tag{4}$$

and the average area ($\overline{A}$) as $\overline{A} = A/n$.

The area A_i of the footprint is a main driver of severity. This can be understood by considering $D_i = L_i/A_i$, the severity per unit area of winds that exceed the threshold. The mean severity $\overline{L} = \overline{AD} = \overline{A}\,\overline{D}\left[1 + v_A v_D\, cor\left(\overline{A}, \overline{D}\right)\right]$, where v denotes the coefficient of variation. For the data in this study, we find that generally $v < 1$ and the correlation is small, and so a good approximation is given by $\overline{L} \approx \overline{A}\,\overline{D}$.

For each of the aforementioned metrics, only the seasonal values (aggregate or average) are considered to allow construction of a time series. Each metric is calculated for each of the analysis regions detailed earlier. Despite each region being assessed individually, windstorms can impact multiple regions.

2.5 | Trend estimation using generalized linear models

For estimating trends in windstorm severity over a full transient time series, we use generalized linear models (GLMs). GLMs are used for this purpose for three reasons, firstly, changes in our time series are unlikely to be linear, and also a GLM estimate will be positive definite. Finally, our data will not be normally distributed; therefore, it is unwise to use a linear model for this. Different distributions and associated link functions are used in the construction of the GLMs for the different metrics. For windstorm counts, we use a Poisson GLM, and for the SSI metrics a Gamma GLM is used, both of which are constructed using a log-link. The GLM assumes that the response variable (windstorm counts/SSI) is distributed following the aforementioned distributions and that the logarithm of its expected value μ can be described by a linear combination of unknown parameters using the predictor t. The GLM is formulated as $\ln(\mu) = \beta_0 + \beta_1 t$.

The outputs of the GLM can be used to quantify trends in the expected value of the distribution. The log-link model can be transformed to $\mu = e^{\beta_0 + \beta_1 t}$.

As the β_1 parameter is associated with the slope of the GLM, this can be expressed as a trend (P) as follows:

$$P = 100\, e^{\beta_1 t} - 100, \tag{5}$$

where t is the unit of change and P is the percentage change in the metric per unit change in t. The most common approach is to use time as the predictor (t) and quantify how the timeseries trends with time. Therefore, a value of $t = 100$ is used to understand the rate of change of our metrics over the course of 100 years.

Instead of regressing on time, the GLM can also be constructed by regressing global mean surface temperature (GMST). Each model has a different evolution of GMST, and changes are often proportional to rate of warming. Trends with time may not be comparable between models, and it is often easier to show changes in model response of hazards and various phenomena to the rate of change of GMST (e.g. Fischer & Knutti, 2016; Harvey et al., 2023; Tabari, 2020). This approach also allows for an increase in model sample size by combining simulations from different scenarios (SSPs). Timeseries of each model GMST are constructed following Harvey et al. (2023), and by using this information in the GLM we can then construct trends per °C warming which would use a value of $t = 1$ in Equation (5).

Model trends from the GLMs are combined using a weighted median approach. The β_1 parameter has a standard deviation associated with that estimate, which is used as a weighting. Models for which there is less confidence in the GLM estimate will have a larger standard deviation of the β_1 parameter, and vice versa. A full description of our weighted median estimator method and confidence interval calculation can be found in the Appendix.

3 | RESULTS

3.1 | Forced trends

3.1.1 | Multimodel assessment

Figure 4 shows the GLM trend estimation method to the aggregate storm severity (L) for the 12 models providing data for the historical and SSP5-85 scenarios in estimates for the WEU region. The GLM (red lines) provides an estimate of the expectation of the distribution in any given year, hence its smooth nature. There is considerable interannual variability in the raw L time series, which is not represented in our GLM trend estimates. The models have a wide range of trends, ranging from +259% per century to −48% per century (Figure 4). For WEU, the weighted median method downweights models such as HadGEM3-GC3.1-MM as the trend estimate is very

difficult to obtain due to the low signal and high variability in this model. The combination of the model estimates can be interpreted as the forced trend as a result of climate change for WEU. Applying our weighted median estimate, a forced trend from the model ensemble is obtained as $+34 \pm 25\%$ (Table 2). As the confidence intervals are smaller than the magnitude of the forced trend, there is high confidence of an increase in L for WEU (see the Appendix for information on confidence interval calculation).

The forced trends in L, are a combination of a number of factors, the first of which is $\overline{L}$, the average SSI per storm, the second is the frequency of windstorms n, and finally the average area of each windstorm footprint ($\overline{A}$). Figure 5 shows the forced trends for each of these metrics, for all regions of Europe, following the methodology demonstrated in Figure 4. The changes noted in L for WEU are also present for GB/IE, with increases in
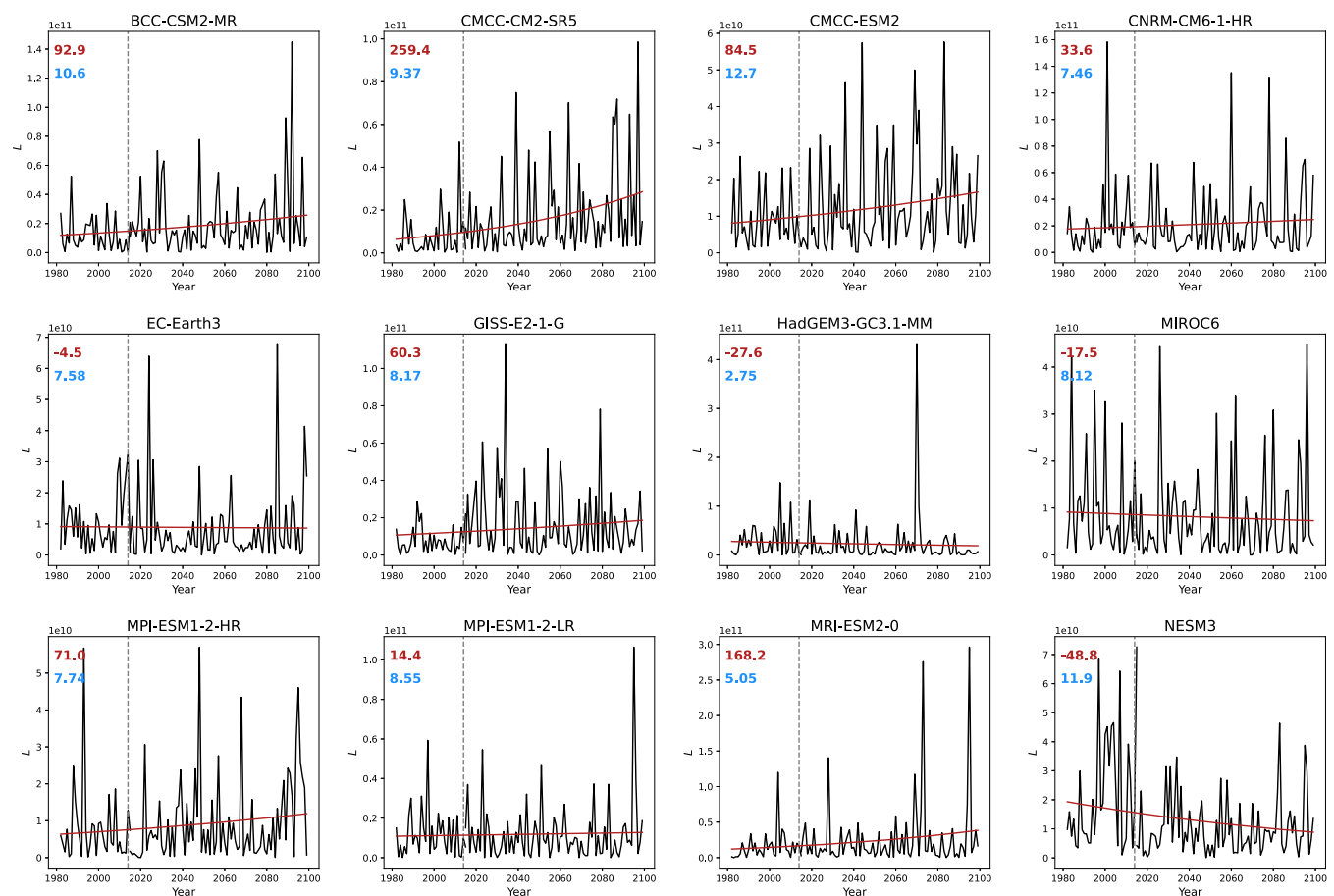
FIGURE 4 Future projections of aggregate storm severity (L) for the western Europe region. Time series of L for the 12 Coupled Model Intercomparison Project Phase 6 (CMIP6) models providing data for the historical and SSP5-85 scenarios. Black lines indicate the L time series from 1980 to 2100. Vertical dashed grey line indicates the concatenation point of the historical and SSP5-85 data. Solid red line is the generalized linear model (GLM) prediction of the expectation of the data from 1980-2100. Red text in the upper left of each panel is the GLM predicted trend in $\Delta\%$ per century. Blue text in the upper left of each panel is the weighting of each model GLM estimate expressed as a percentage.

TABLE 2 Table of windstorm metric changes from historical and SSP5-85 data from 1980 to 2100. For each metric the stated change is the weighted median change with two standard errors of that trend estimate. All trends are in units of $\Delta\%$ per century.

Region	L	$\overline{L}$	n	$\overline{A}$
SC	-33.0 ± 23.7	-13.4 ± 21.4	-9.6 ± 5.0	-9.2 ± 6.7
GB/IE	31.3 ± 27.7	21.7 ± 25.2	-0.8 ± 6.8	0.4 ± 5.4
WEU	33.6 ± 24.8	27.2 ± 23.1	1.5 ± 6.1	4.7 ± 5.6
CEU	12.3 ± 26.5	15.3 ± 24.8	-2.6 ± 6.6	2.0 ± 6.4
EEU	13.1 ± 22.0	13.8 ± 20.4	-12.0 ± 5.4	0.7 ± 5.9
IB	-25.3 ± 30.9	-4.9 ± 30.5	-7.2 ± 6.6	-9.3 ± 7.0
MED	-8.5 ± 21.0	0.5 ± 19.9	-12.4 ± 5.5	-0.7 ± 5.4
EU	-6.0 ± 13.7	13.1 ± 13.2	-8.4 ± 3.9	-3.6 ± 4.9

CEU: Austria, Czechia, Germany, Switzerland; EEU: Belarus, Hungary, Poland, Republic of Moldova, Romania, Slovakia, Ukraine; EU: all Europe; GB/IE: Great Britain, Ireland; IB: Spain, Portugal; MED: Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Italy, Greece, Macedonia, Malta, Montenegro, San Marino, Serbia, Slovenia; SC: Denmark, Estonia, Finland, Latvia, Lithuania, Norway, Sweden; WEU: Belgium, France, Luxembourg, Netherlands.

aggregate severity of $+31(\pm28)\%$ per century (Figure 5a). Declines are noted for northern (SC) and southern (IB, MED) Europe. There are increases for CEU and EEU; however, the confidence intervals on these estimates are larger than the weighted median estimate. The specific changes and confidence intervals for these regions are shown in Table 2. Changes in storm average severity ($\overline{L}$) (Figure 5b) are consistent with L (Figure 5a), with increases for WEU and GB/IE and the WEU changes having high confidence, and declines or low confidence changes for the rest of Europe.

For trends in the frequency of windstorms n (Figure 5c) there is an increase in frequency for WEU, with all other regions showing a weighted median decline of up to -12% per century, with this being most extreme, and having high confidence, for MED, IB, EEU, and SC. The increase in windstorm frequency noted for WEU (Figure 5c) is consistent with previous studies (e.g. Little *et al.*, 2023; Priestley & Catto, 2022a; Zappa *et al.*, 2013). However, there is large variability in our n estimate for this regions, with our weighted median estimates having confidence intervals larger than the forced trend, indicating low confidence (see Table 2). This is also the case for GB/IE and CEU, which indicate a weighted median decrease, yet the confidence intervals are considerably larger.

Figure 5d shows the change in average windstorm area $\overline{A}$. This has a pattern of change similar to that in Figure 5a,b, albeit with lower magnitudes. Therefore, we find that increases in aggregate storm severity in GB/IE

and WEU is a result of an increase in average storm intensity and the area of the windstorm footprint, despite an uncertain sign of change in frequency.

There are certain regions/metrics where confidence in forced-trend estimates is higher. This confidence is determined by the size of the confidence intervals relative to the forced trends (Table 2). Confidence is high for forced trends of windstorm counts in southern Europe (IB, MED), however the severity metrics indicate low confidence. Confidence is also high of a decline in n and L for SC, but changes in $\overline{L}$ are less certain, and have low confidence. Confidence is high for an increase in L for GB/IE and WEU, although changes in n have low confidence. For $\overline{L}$, GB/IE and WEU trends have medium and high confidence respectively. CEU and EEU exhibit the lowest confidence overall, with this being especially the case for CEU. For these regions, all metrics have confidence intervals larger than the forced trends, with only the decline in n for EEU having a clear signal. Therefore, under the SSP5-85 scenario, any projections in windstorm severity for these regions should be taken with caution.

3.1.2 | Changes per °C warming

By regressing on GMST in our GLMs, we can use all 20 CMIP6 simulations across SSP2-45 and SSP5-85 in the weighted median trend estimates. These are shown in Figure 6 with trends expressed as percentage change per °C global mean warming.

The pattern of trends seen per °C warming (Figure 6) are broadly consistent with those seen for just the SSP5-85 scenario models (Figure 5). Where there have been changes in sign of trend, there is uncertainty on the sign of the trend for that region. Increases in L and $\overline{L}$ are seen for GB/IE, WEU, CEU, and EEU, with decreases for IB, MED, and SC. The weighted trend increases for GB/IE and WEU in L are by $1.1\%\cdot{}^{\circ}C^{-1}$ and $14.2\%\cdot{}^{\circ}C^{-1}$ respectively and in $\overline{L}$ by $2.9\%\cdot{}^{\circ}C^{-1}$ and $8.7\%\cdot{}^{\circ}C^{-1}$ respectively, and continue to have high confidence for WEU. For EEU, where changes were uncertain in Figure 5, there continues to be considerable uncertainty (Table 3). The previously uncertain weighted median trend increases for CEU now have high confidence in both L and $\overline{L}$ (Figure 6a,b) of $11.3\%\cdot{}^{\circ}C^{-1}$ and $6.9\%\cdot{}^{\circ}C^{-1}$ respectively. Declining trends in L and $\overline{L}$ are seen for SC, IB, and MED of $-6\%\cdot{}^{\circ}C^{-1}$, $-12.1\%\cdot{}^{\circ}C^{-1}$, and $-3\%\cdot{}^{\circ}C^{-1}$ and $-5.9\%\cdot{}^{\circ}C^{-1}$, $-2.1\cdot{}^{\circ}C^{-1}$, and $-1.4\%\cdot{}^{\circ}C^{-1}$ respectively.

With regard to n and $\overline{A}$ (Figure 6c,d), the sign of forced trends continues to be consistent with Figure 5c,d for most regions. A trend in n is seen for all regions of up to $-3.7\%\cdot{}^{\circ}C^{-1}$, with the sign of this trend having low
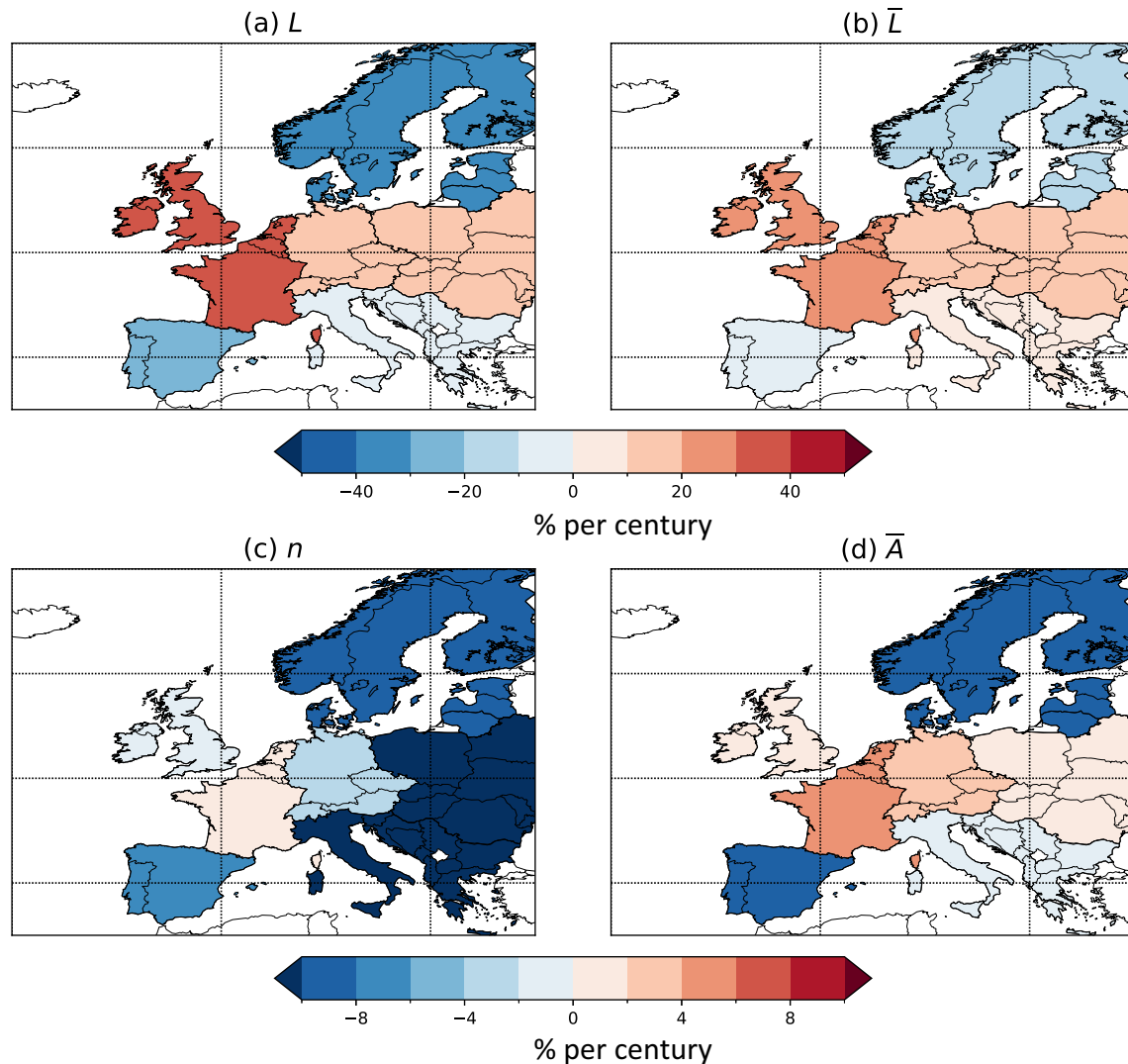
FIGURE 5 Weighted median trend estimates of European windstorm severity metrics for different regions of Europe: (a) the seasonal aggregate severity (L); (b) the storm average severity ($\overline{L}$); (c) the frequency of windstorms (n); (d) the storm average footprint area ($\overline{A}$). All units are Δ% per century. Panels (a) and (b) extend to ±50% per century, and panels (c) and (d) extend to ±10% per century.

confidence for GB/IE and WEU. For $\overline{A}$, a low confidence increase is noted for WEU and CEU of up to $+1.3\% \cdot °C^{-1}$, with reductions of up to $-3.3\% \cdot °C^{-1}$ for the remaining regions. Magnitudes of changes in Figure 6 are of course smaller than in Figure 5 due to the standardizations of trends by the rate of GMST; however, the inclusion of SSP2-45 experiments has resulted in the pattern of change seen in our weighted median trend estimates being largely unchanged.

The addition of the SSP2-45 scenario data has increased the uncertainty in the forced trend from climate change for some metrics/regions. The most significant is for L for GB/IE. These now have confidence intervals larger than the forced trend, indicating lower confidence (Table 3). This potentially indicates that the forced trends may be sensitive to the rate of forcing or magnitude of

total GMST change. Some regional trends, such as for frequency in northern and southern Europe (SC, MED, IB), continue to exhibit very confident decreases in the forced trends per °C. Confidence of a decline in windstorm frequency for these regions is therefore very high.

3.1.3 | Joint trends in frequency and severity

By separating out the aggregate SSI into components from average severity, frequency, and area (Figures 5 and 6), the contributions to the aggregate change from each component have been quantified. This allows for relationships between these different components to be established. Changes in windstorm average area dominate changes in average severity, and this is clearly evident in Figure 7c,f
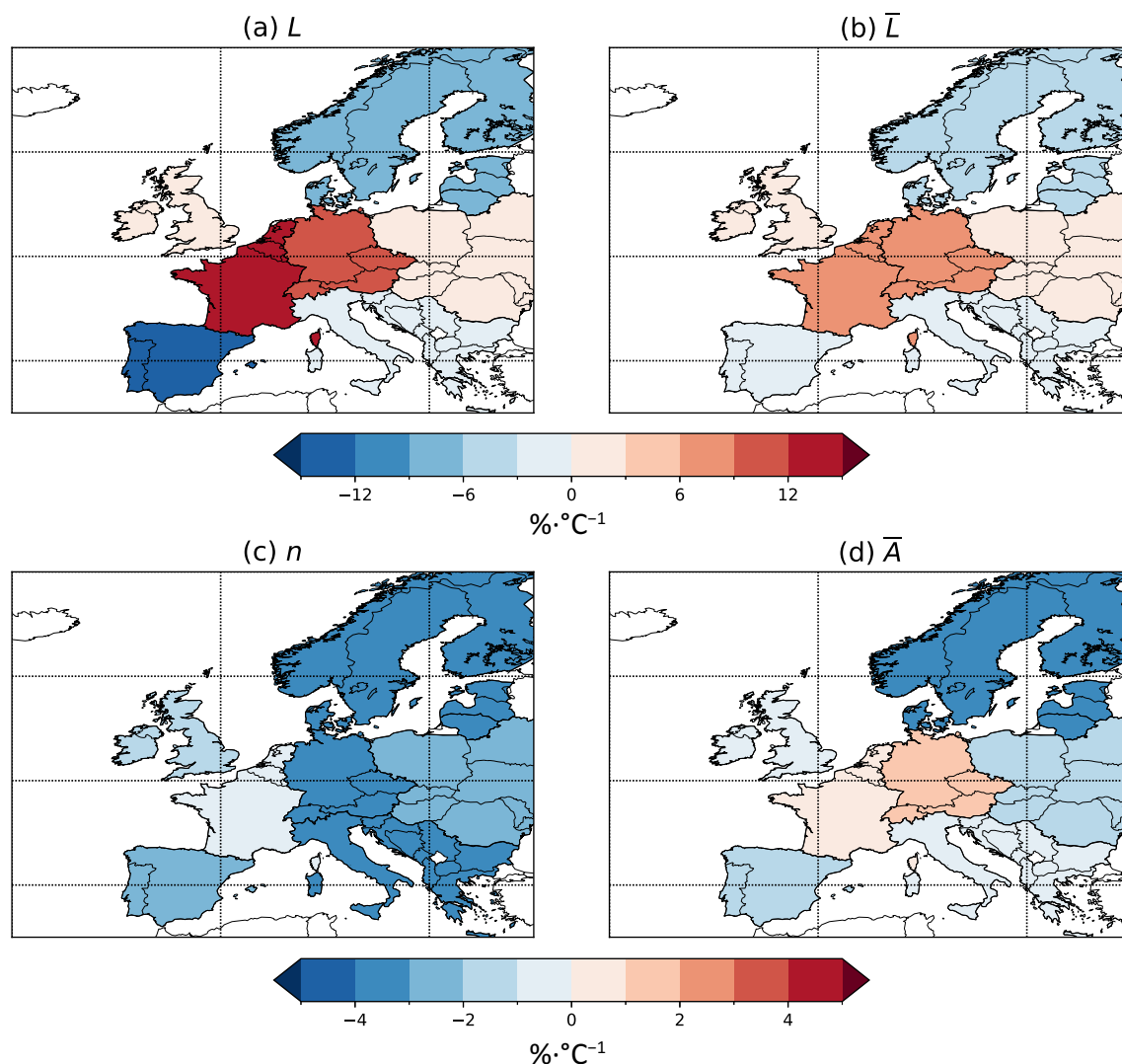
FIGURE 6 Weighted median trend estimates of European windstorm severity metrics for different regions of Europe using global mean surface temperature as our generalized linear model predictor: (a) the seasonal aggregate severity (L); (b) the storm average severity ($\overline{L}$); (c) the frequency of windstorms (n); (d) the storm average footprint area ($\overline{A}$). All units are $\Delta\%\cdot°C^{-1}$. Panels (a) and (b) extend to $\pm15\%\cdot°C^{-1}$, and panels (c) and (d) extend to $\pm5\%\cdot°C^{-1}$.

(r=0.91, see also Priestley & Catto, 2022a). Figure 7 also shows a relationship between frequency of windstorms and the average severity; however, the area of a footprint is most strongly associated with variations in the mean severity. This is evident for both the trends calculated per century (Figure 7a–c) and per °C warming (Figure 7d–f).

3.2 | Quantifying model and total variability

The choice of model is not the only factor influencing variability in projections; another consideration is that of internal variability, and hence the choice of ensemble member used for analysis. Utilizing a 15-member

ensemble from the MPI-ESM1-2-LR model, the role of internal variability on our European storminess metrics can be quantified. Furthermore, the internal variability from the MPI-ESM1-2-LR ensemble can be compared in magnitude with that of the already calculated variability from the independent multimodel ensemble from CMIP6, both with SSP5-85 forcing.

As MPI-ESM1-2-LR is the only model to provide more than one ensemble member (Table 1) with the necessary variables required for our analysis, we assume that its internal variability is representative of the other models. To quantify the internal and total variability, the CMIP6 model and MPI-ESM1-2-LR member trends are independently bootstrapped to create a 50-member ensemble, with the forced trend being estimated from this

TABLE 3 Table of windstorm metric changes from historical, SSP2-45, and SSP5-85 data from 1980 to 2100 using global mean surface temperature as our generalized linear model predictor. For each metric the stated change is the weighted median change with two standard errors of that trend estimate. All trends are in units are $\Delta\% \cdot {}^{\circ}C^{-1}$.

Region	L	$\overline{L}$	n	$\overline{A}$
SC	-6.0 ± 5.2	-5.9 ± 4.7	-3.7 ± 1.2	-3.3 ± 1.5
GB/IE	1.1 ± 6.2	2.9 ± 5.7	-1.4 ± 1.6	-0.1 ± 1.3
WEU	14.2 ± 5.6	8.7 ± 5.2	-0.7 ± 1.4	0.3 ± 1.3
CEU	11.3 ± 6.0	6.9 ± 5.6	-3.6 ± 1.5	1.3 ± 1.5
EEU	0.0 ± 4.9	1.5 ± 4.6	-2.1 ± 1.3	-1.3 ± 1.4
IB	-12.1 ± 6.8	-2.1 ± 6.7	-2.5 ± 1.6	-1.9 ± 1.6
MED	-3.0 ± 4.7	-1.4 ± 4.4	-3.7 ± 1.3	-0.8 ± 1.3
EU	-3.4 ± 3.1	1.0 ± 3.0	-2.8 ± 0.9	-0.9 ± 1.1

CEU: Austria, Czechia, Germany, Switzerland; EEU: Belarus, Hungary, Poland, Republic of Moldova, Romania, Slovakia, Ukraine; EU: all Europe; GB/IE: Great Britain, Ireland; IB: Spain, Portugal; MED: Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Italy, Greece, Macedonia, Malta, Montenegro, San Marino, Serbia, Slovenia; SC: Denmark, Estonia, Finland, Latvia, Lithuania, Norway, Sweden; WEU: Belgium, France, Luxembourg, Netherlands.

via the median. This is done 10,000 times, and the spread (standard deviation σ) of these 10,000 trend estimates is assessed. This spread represents the role of internal variability, and total variability, on uncertainty in the respective forced-trend estimates. The total variability is calculated from the multimodel ensemble as it comprises of differences between the models and internal variability across these simulations. If internal variability is small relative to total variability, the spread in the 10,000 forced-trend estimates will be small and vice versa.

Figure 8 shows the spread of the 10,000 bootstrapped samples for the CMIP6 models, and the MPI-ESM1-2-LR ensemble. Calculating the uncertainty from internal variability relative to total variability (Figure 8i–l), it is apparent that internal variability is always smaller than total variability for GB/IE, WEU, and SC, with uncertainty from internal variability never being more than 80% of the uncertainty from total variability, and commonly less than 50% in magnitude. For aggregate severity (L) (Figure 7i), internal variability is less than total variability for all regions except for MED and is only 30–60% in magnitude for the regions of GB/IE, WEU, CEU, and SC. For certain regions/metrics, the internal variability is very large (e.g.,
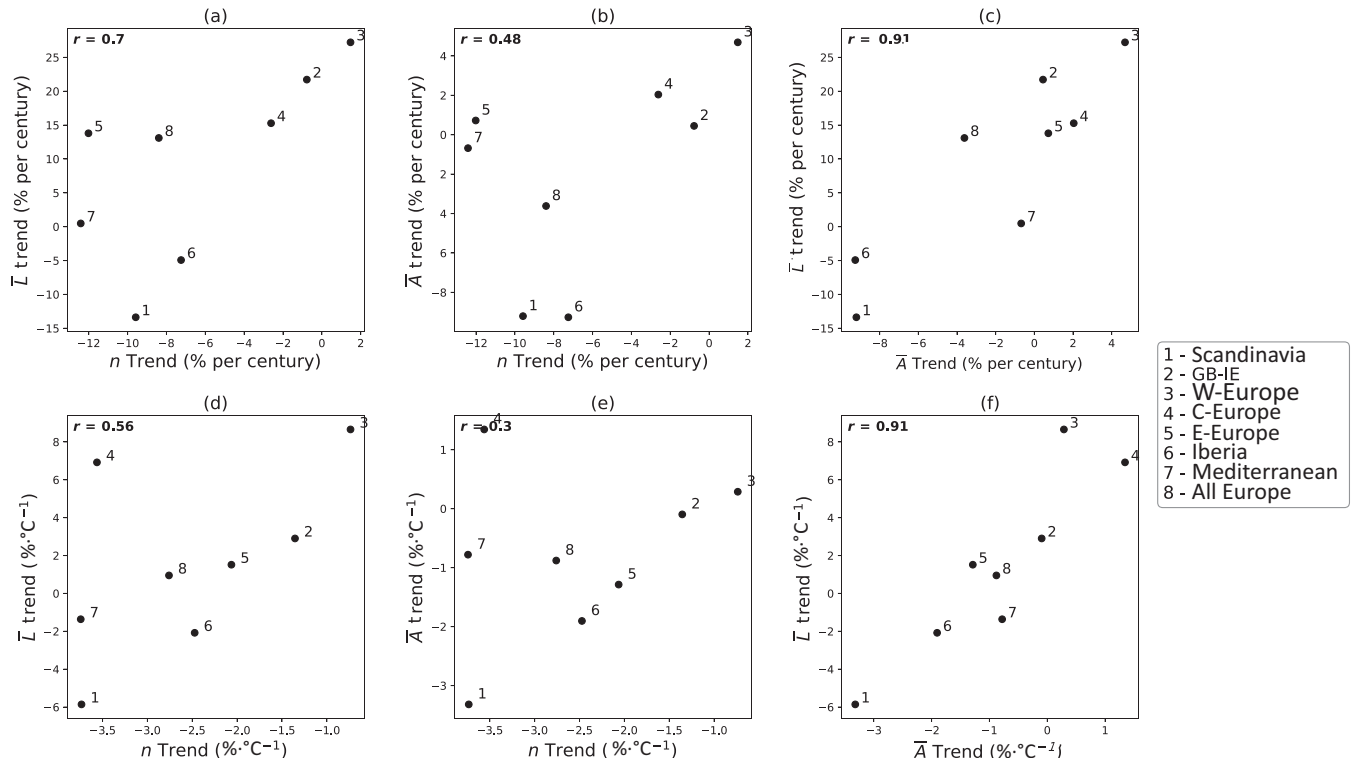
FIGURE 7 Relationship between windstorm metrics. Scatter plots of the weighted median trends in different windstorm severity metrics for each subregion of Europe: (a–c) trends per century; (d–f) trends per °C warming. (a, d) Storm average severity ($\overline{L}$) against frequency (n). (b, e) Average windstorm area ($\overline{A}$) against frequency (n). (c, f) Storm average severity ($\overline{L}$) against average windstorm area ($\overline{A}$). All units are $\% \cdot {}^{\circ}C^{-1}$. Text in the upper left of each panel indicates the correlation coefficient of the seven core analysis regions (points 1–7).

IB and CEU in Figure 8k,l). Therefore, when examining trends in windstorm severity, consideration must be made as to the uncertainty from internal variability present when performing analysis with only a single realization. It should be noted that the total variability (Figure 8a–d) is potentially reducible, whereas the uncertainty from internal variability (Figure 8e–h) is not. This may be possible via development of improved models, thereby resulting in the internal variability being a much larger fraction of the total variability than noted here.

Examining the variability in each metric, the spread is largest for L, followed by $\overline{L}$, in both the CMIP6 and MPI ensembles, with variability being reduced further still for n and $\overline{A}$. Variability in forced-trend estimates tends to be largest for the central and northern latitudes of Europe. In the CMIP models, the spread in forced-trend estimates

tends to be 15–25% (5–15% for MPI) for L and $\overline{L}$ for all regions apart from MED, indicating that even when resampling the models/ensemble members to capture more variability that estimates in the forced trends can still be very different. For n and $\overline{A}$ (Figure 8c,d,g,h), uncertainty in the forced trends is much smaller, and generally less than 5%, which agrees with the smaller projected trends of these metrics (Figure 5c,d).

Despite internal variability being lower than total variability in estimating the forced trend, the trends between individual members can have large variations. Some ensemble members have similar trends for L, but there is a large difference between the members with the extreme positive and negative trends. The standard deviation of the trends from the 15 ensemble members is shown in Figure 9. The largest variability in trends is in L and $\overline{L}$
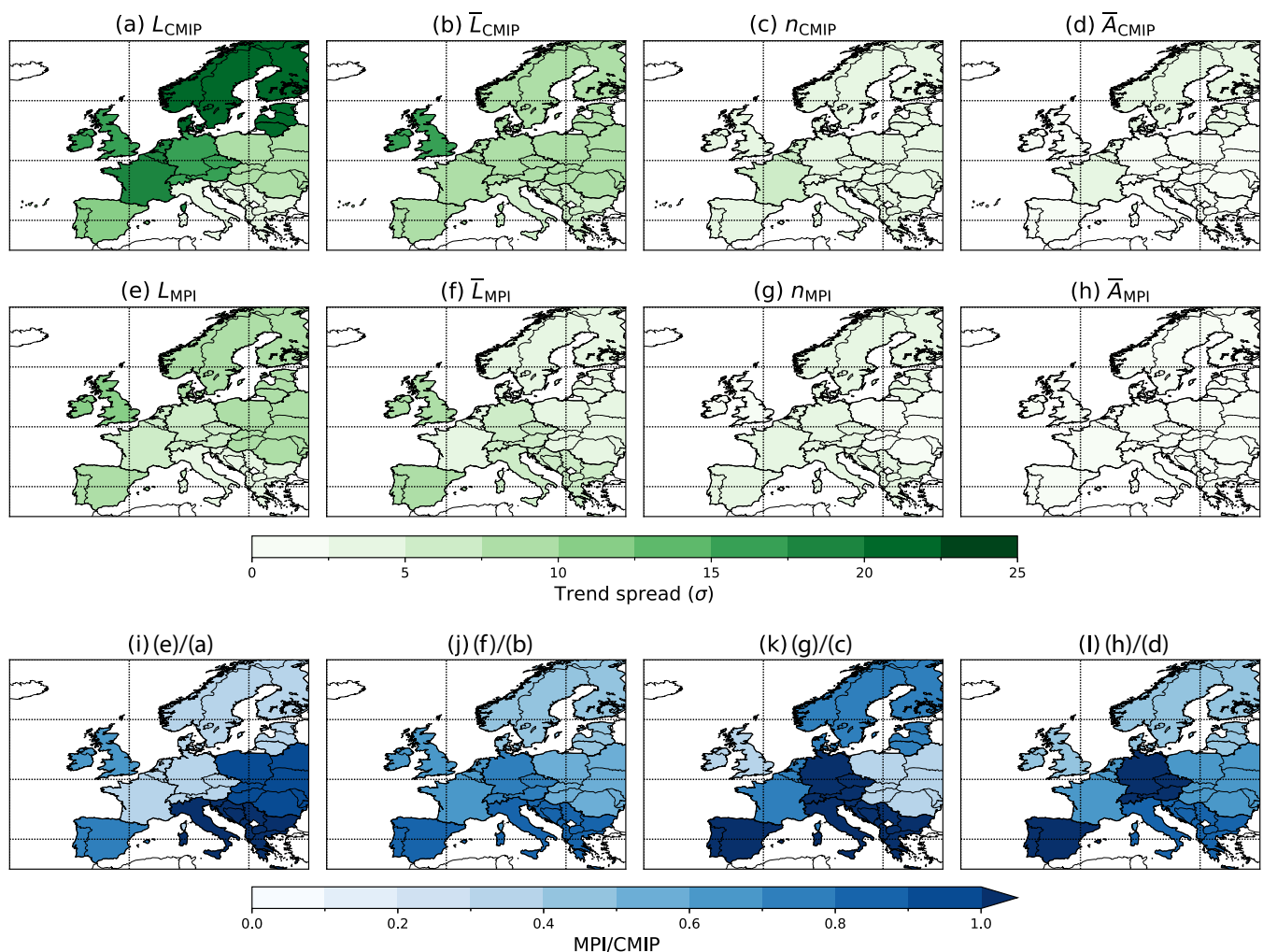
FIGURE 8   Relative importance of internal variability compared with total variability in windstorm trends: (a, e) the seasonal aggregate severity (L); (b, f) the storm average severity ($\overline{L}$); (c, g) the frequency of windstorms (n); (d, h) the storm average footprint area ($\overline{A}$). (a–d) Total variability from the CMIP6 multimodel ensemble. (e–h) Internal variability from MPI-ESM1-2-LR ensemble. Bottom row shows the size of internal variability relative to total variability. Spread is represented as the standard deviations (σ) of the 10,000 bootstrap sample estimates of the trend. Units of the top two rows are %.
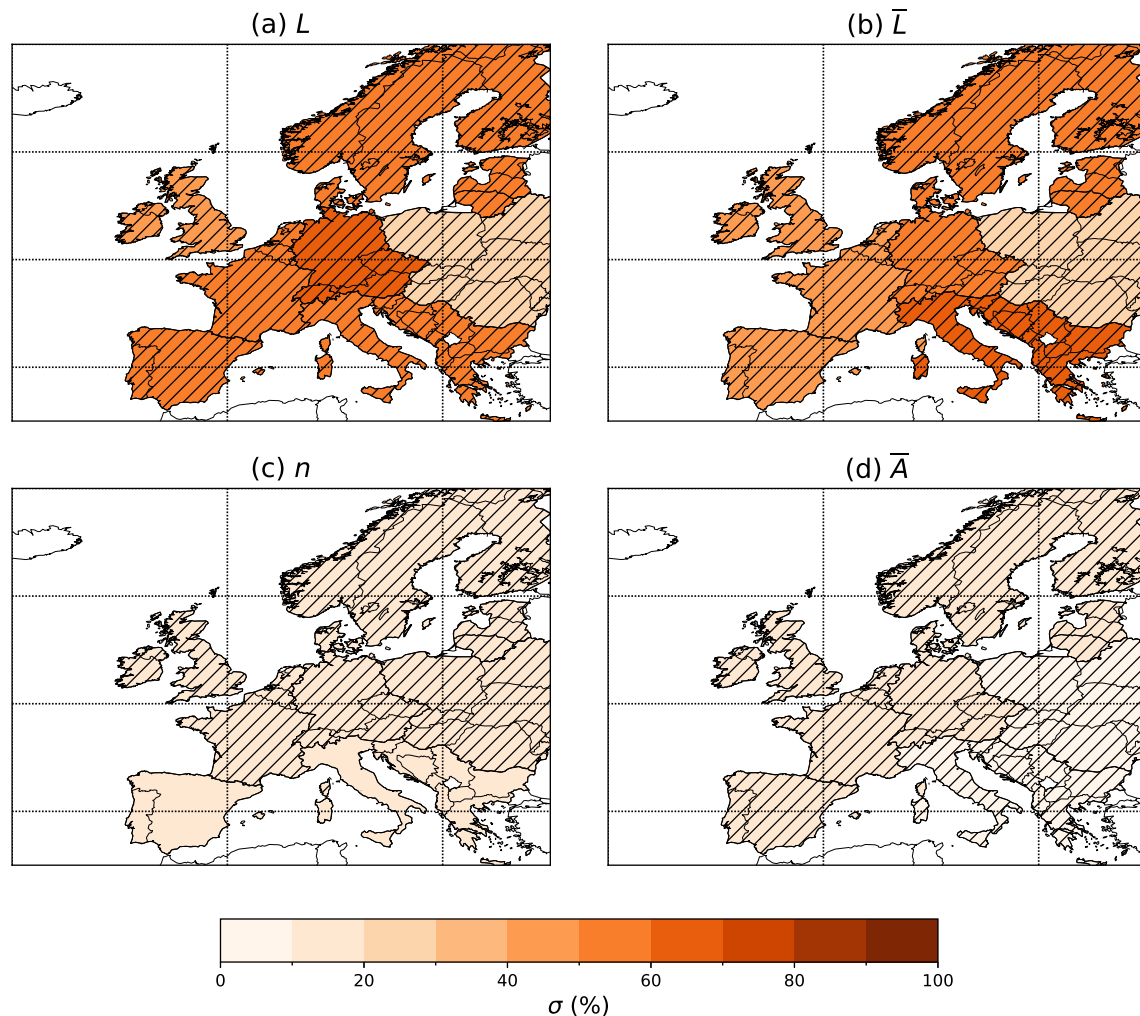
FIGURE 9 Standard deviation of MPI-ESM1-2-LR ensemble member projected trends from 1980 to 2100 using the historical and SSP5-85 scenarios: (a) the seasonal aggregate severity (L); (b) the storm average severity ($\overline{L}$); (c) the frequency of windstorms (n); (d) the storm average footprint area ($\overline{A}$). All units are per cent. Hatching indicates where the standard deviation is larger than the median ensemble member trend.

(Figure 9a,b), which is a result of these metrics having the largest magnitudes. For L and $\overline{L}$, all regions have a standard deviation of the trends that is larger than the median estimate. The standard deviations are largest for CEU, and this is by up to 70%. The lowest variability is in the EEU region. For n (Figure 9c), the standard deviation of trends is generally small; however, there is only consistency between the members for MED and IB. Finally, for $\overline{A}$ (≤10%) there is limited consensus from the ensemble members on the sign of this change (Figure 9d). Therefore, even though internal variability is generally smaller than total variability, it is clear that substantial differences can arise from one member to the next, thereby demonstrating a need for large ensemble analyses of European windstorm severity in future modelling experiments.

4 | CONCLUSIONS

This study has provided an estimate of the forced trend from climate change of European windstorm metrics from an ensemble of CMIP6 projections, and also quantified the impact of internal variability. This work combines a quantification of windstorm footprints associated with ETCs, with a statistical analysis of forced trends via a weighted median, to estimate future trends in windstorm severity and frequency. Aggregate severity has been separated into contributions from average windstorm severity, windstorm frequency, and also windstorm footprint area. The main finding is that storm severity metrics (both intensity and area) show increasing trends over the central latitudes of Europe

(GB/IE, WEU, CEU), with reductions for southern Europe and Scandinavia. Decreases in windstorm frequency were found across all of Europe. The internal variability of windstorm trends was quantified, with the spread between ensemble members of MPI-ESM1-2-LR often being much larger than the median change of the members. However, the uncertainty from internal variability is generally smaller than the total variability, and hence the uncertainty due to model formulation.

The multimodel assessment of a number of different metrics allows a quantification of confidence in a number of the findings by region as follows:

1. *High*
   - Increases in storm aggregate and average severity for WEU.
   - Decreases in aggregate severity, windstorm frequency, and footprint area for northern Europe (SC).
   - Decreases in windstorm frequency for southern and eastern Europe (MED, IB, EEU).

2. *Medium*
   - Increases in aggregate and storm average severity for GB/IE and CEU.
   - Decreases in average storm severity northern Europe (SC).
   - Decreases in aggregate severity for IB.

3. *Low*
   - Decreases in storm frequency and changes in footprint area for northwestern Europe (GB/IE, WEU, CEU).
   - Decreases in storm average severity for southern and eastern Europe (MED, IB, EEU).
   - Decreases in aggregate severity for MED and EEU.

Confidence in our findings is determined by a number of factors. First, the magnitude of the weighted median estimates presented in Figures 5 and 6, because those with larger magnitudes are likely to be due to a stronger signal. It is also based on the confidence intervals shown in Tables 2 and 3, with those regions and metrics that have larger confidence intervals relative to their weighted medians having lower confidence, an example of this being trends in $\overline{L}$ for MED and IB. A final factor is when there is a deviation, or change, in either of these metrics when examining the SSP5-85 experiments, or the combination of SSP2-45 and SSP5-85 experiments with changes per °C. An example is the trends in L and $\overline{L}$ for GB/IE, as both have increasing trends in Figures 5 and 6; however, confidence intervals are small in Table 2 but are much larger and cover zero in Table 3. This suggests that trends

may only be apparent in the highest forcing scenario; therefore, when examining just changes in the SSP5-85 scenario a confidence label of *High* would be assigned, but with the uncertainty that arises when including lower forcing scenarios we can only assign *Medium* confidence to these trends.

These findings offer increased robustness over previous studies (e.g. Barcikowska *et al.*, 2018; Donat *et al.*, 2011; Little *et al.*, 2023; Pinto *et al.*, 2007; Schwierz *et al.*, 2010; Severino *et al.*, 2024) for several key reasons. First, windstorm footprints are associated with ETCs, which are the known drivers of high winds over Europe in winter. Therefore, other wind events, such as those that may be convective driven, are not considered (Severino *et al.*, 2024). Also, windstorm footprints that cover all of Europe are assessed (see Figure 1a). Even though ETCs have a primary influence of 5–10° from an objective centre (Little *et al.*, 2023), the impact may extend to much greater distances, which the method presented here is able to capture. Trends are also assessed across a continuous time series from 1980 to 2100, whereas most studies often approach climate change studies through a time-slice approach (e.g. Leckebusch *et al.*, 2007; Priestley & Catto, 2022a; Schwierz *et al.*, 2010). This approach ensures that any changes identified are less sensitive to different representations of interannual and multi-decadal variability and more representative of each model's trend throughout the 20th and 21st centuries.

Trends are quantified using weighted trend estimates, which is a way to combine a multimodel ensemble that goes beyond analysing model means or the equal weighting of models (Knutti *et al.*, 2010). It has been demonstrated that many models have differing representation of the European climate (Palmer *et al.*, 2023), and that models which do not have a good historical representation likely do not accurately capture the climate response (Whetton *et al.*, 2007). This weighted median estimate moves toward this without the need for an assessment of model fidelity, as it is likely that models with lower weightings will be a result of biases in that model's representation of numerous large-scale features, and vice versa.

The variability in trends between single ensemble members of the MPI-ESM1-2-LR model and the individual CMIP6 GCMs highlights that single-model, single-member studies should be avoided for investigations of this nature. With continued advances in modelling of the climate system, it is likely that model variability will reduce, whereas internal variability is likely to remain constant. In future assessments, it may be that more consideration needs to be taken on the choice of ensemble size rather than the choice of GCM. Even when resampling to 50 members there is still variability in our trend estimates of 10%–15%, indicating that larger ensembles are required to accurately estimate the forced trend. This study has not

discussed the role of interannual variability as our method is based on the expectation of the distribution, and therefore ignores year-to-year variability. Interannual variations are notably larger than the trend estimates (see Figure 4); as a result, those that may be financially implicated by high loss events such as European windstorms will need to consider these variations at least as carefully as the trend estimates presented in this study.

There are several caveats to this work and its findings:

- Storm severity metrics, even though widely adopted, are not representative of actual losses. This would involve either simulation through a full catastrophe model or coupling to a loss model such as CLIMADA—as adopted in (as adopted in Severino *et al.*, 2024). However, we hypothesize that our findings are representative of the changes expected in insured losses due to the high correlation between SSI and such losses (Klawa & Ulbrich, 2003).

- Only 13 unique models across two SSPs, totalling 20 total realizations, have been analysed. This is likely to be representative of the majority of CMIP6 model variability; however, increasing the number of GCMs, and ensemble members, will always be preferential.

- The weighted median assessment is designed to be a simplistic step from equal weighting of a model ensemble toward a filtering based on model fidelity. It is an idealised method of doing this and future work should be directed toward a systematic subsetting and physical analysis of model trends to generate an ensemble based upon those that have the most realistic representation of historical variability and trends.

- It may be that the forced signal in European storminess may be considerably underrepresented (the signal-to-noise paradox; Scaife & Smith, 2018). The true magnitude of the trend may only be realized with improved modelling capabilities that address factors such as improving the underresolved eddy feedbacks (Hardiman *et al.*, 2022; Scaife *et al.*, 2019) that may strengthen teleconnections that are currently poorly represented, such as that between El Niño–Southern Oscillation and the North Atlantic Oscillation (Williams *et al.*, 2023).

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

Code is available at the request of the author. CMIP6 data are publicly available through the Earth System Grid Federation. The cyclone tracking and compositing algorithm TRACK is available at the request of Kevin Hodges from https://gitlab.act.reading.ac.uk/track/track (Hodges, 1994, 1995, 1999).

ORCID

Matthew D. K. Priestley https://orcid.org/0000-0002-5488-3959

REFERENCES

Barcikowska, M.J., Weaver, S.J., Feser, F., Russo, S., Schenk, F., Stone, D.A. et al. (2018) Euro-Atlantic winter storminess and precipitation extremes under 1.5 °C vs. 2 °C warming scenarios. *Earth System Dynamics*, 9, 679–699.

Barredo, J.I. (2010) No upward trend in normalised windstorm losses in Europe: 1970–2008. *Natural Hazards and Earth System Sciences*, 10, 97–104.

Bechtold, P. & Bidlot, J. (2009) Parametrization of convective gusts. *ECMWF Newsl*, 199, 15–18.

Bengtsson, L., Hodges, K.I. & Roeckner, E. (2006) Storm Tracks and Climate Change. *Journal of Climate*, 19, 3518–3543.

Blanusa, M.L., López-Zurita, C.J. & Rasp, S. (2023) Internal variability plays a dominant role in global climate projections of temperature and precipitation extremes. *Climate Dynamics*, 61, 1931–1945.

Bloomfield, H., Hillier, J., Griffin, A., Kay, A., Shaffrey, L., Pianosi, F. et al. (2023) Co-occurring wintertime flooding and extreme wind over Europe, from daily to seasonal timescales. *Weather and Climate Extremes*, 39, 100550.

Born, K., Ludwig, P. & Pinto, J.G. (2012) Wind gust estimation for Mid-European winter storms: towards a probabilistic view. *Tellus A: Dynamic Meteorology and Oceanography*, 64, 17471.

Brehm, W. (2023) *A Holistic Approach to Structural Dynamics Using Serial Crystallography*. Ph.D. thesis, University of Hamburg. Available from: https://ediss.sub.uni-hamburg.de/handle/ediss/10656

Carvalho, D., Rocha, A., Costoya, X., deCastro, M. & Gómez-Gesteira, M. (2021) Wind energy resource over Europe under CMIP6 future climate projections: What changes from CMIP5 to CMIP6. *Renewable and Sustainable Energy Reviews*, 151, 111594.

Catto, J.L., Shaffrey, L.C. & Hodges, K.I. (2011) Northern hemisphere extratropical cyclones in a warming climate in the HiGEM high-resolution climate model. *Journal of Climate*, 24, 5336–5352.

Clark, P.A. & Gray, S.L. (2018) Sting jets in extratropical cyclones: A review. *Quarterly Journal of the Royal Meteorological Society*, 144, 943–969.

Cusack, S. (2023) A long record of European windstorm losses and its comparison to standard climate indices. *Natural Hazards and Earth System Sciences*, 23, 2841–2856.

Deser, C., Hurrell, J.W. & Phillips, A.S. (2017) The role of the North Atlantic Oscillation in European climate projections. *Climate Dynamics*, 49, 3141–3157.

Deser, C., Phillips, A., Bourdette, V. & Teng, H. (2012) Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 38, 527–546.

Deser, C., Phillips, A.S., Alexander, M.A. & Smoliak, B.V. (2014) Projecting North American Climate over the next 50 years: Uncertainty due to Internal Variability. *Journal of Climate*, 27, 2271–2296.

Donat, M.G., Leckebusch, G.C., Wild, S. & Ulbrich, U. (2011) Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations. *Natural Hazards and Earth System Sciences*, 11, 1351–1370.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. et al. (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9, 1937–1958.

Feser, F., Barcikowska, M., Krueger, O., Schenk, F., Weisse, R. & Xia, L. (2015) Storminess over the North Atlantic and northwestern Europe—A review. *Quarterly Journal of the Royal Meteorological Society*, 141, 350–382.

Fischer, E.M. & Knutti, R. (2016) Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, 6, 986–991.

Geng, Q. & Sugi, M. (2003) Possible change of extratropical cyclone activity due to enhanced greenhouse gases and sulfate aerosols—Study with a high-resolution AGCM. *Journal of Climate*, 16, 2262–2274.

Hardiman, S.C., Dunstone, N.J., Scaife, A.A., Smith, D.M., Comer, R., Nie, Y. et al. (2022) Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions. *npj Climate and Atmospheric Science*, 5, 57.

Harvey, B., Hawkins, E. & Sutton, R. (2023) Storylines for future changes of the North Atlantic jet and associated impacts on the UK. *International Journal of Climatology*, 43, 4424–4441.

Harvey, B.J., Cook, P., Shaffrey, L.C. & Schiemann, R. (2020) The response of the northern hemisphere storm tracks and jet streams to climate change in the CMIP3, CMIP5, and CMIP6 climate models. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032701.

Hawkins, E. & Sutton, R. (2009) The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90, 1095–1108.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Hodges, K.I. (1994) A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, 122, 2573–2586.

Hodges, K.I. (1995) Feature tracking on the unit sphere. *Monthly Weather Review*, 123, 3458–3465.

Hodges, K.I. (1999) Adaptive constraints for feature tracking. *Monthly Weather Review*, 127, 1362–1373.

Karremann, M.K., Pinto, J.G., Reyers, M. & Klawa, M. (2014) Return periods of losses associated with European windstorm series in a changing climate. *Environmental Research Letters*, 9, 124016.

Klawa, M. & Ulbrich, U. (2003) A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Sciences*, 3, 725–732.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G.A. (2010) Challenges in combining projections from multiple climate models. *Journal of Climate*, 23, 2739–2758.

Leckebusch, G.C., Ulbrich, U., Fröhlich, L. & Pinto, J.G. (2007) Property loss potentials for european midlatitude storms in a changing climate. *Geophysical Research Letters*, 34, L05703.

Little, A.S., Priestley, M.D.K. & Catto, J.L. (2023) Future increased risk from extratropical windstorms in northern Europe. *Nature Communications*, 14, 4434.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R. et al. (2019) Developments in the MPI-M earth system model version 1.2 (MPI-ESM1.2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, 11, 998–1038.

Minola, L., Zhang, F., Azorin-Molina, C., Pirooz, A.A.S., Flay, R.G.J., Hersbach, H. et al. (2020) Near-surface mean and gust wind speeds in ERA5 across Sweden: towards an improved gust parametrization. *Climate Dynamics*, 55, 887–907.

Mizuta, R., Matsueda, M., Endo, H. & Yukimoto, S. (2011) Future change in extratropical cyclones associated with change in the upper troposphere. *Journal of Climate*, 24, 6456–6470.

Mölter, T., Schindler, D., Albrecht, A.T. & Kohnle, U. (2016) Review on the projections of future storminess over the North Atlantic European region. *Atmosphere*, 7, 1–40.

Neu, U., Akperov, M.G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R. et al. (2013) IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, 94, 529–547.

O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K.L., Hallegatte, S., Carter, T.R. et al. (2014) A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, 122, 387–400.

O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G. et al. (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9, 3461–3482.

Palmer, T.E., McSweeney, C.F., Booth, B.B.B., Priestley, M.D.K., Davini, P., Brunner, L. et al. (2023) Performance-based sub-selection of CMIP6 models for impact assessments in Europe. *Earth System Dynamics*, 14, 457–483.

Panofsky, H.A., Tennekes, H., Lenschow, D.H. & Wyngaard, J.C. (1977) The characteristics of turbulent velocity components in the surface layer under convective conditions. *Boundary-Layer Meteorology*, 11, 355–361.

Pinto, J.G., Fröhlich, E.L., Leckebusch, G.C. & Ulbrich, U. (2007) Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM. *Natural Hazards and Earth System Sciences*, 7, 165–175.

Priestley, M.D.K. & Catto, J.L. (2022a) Future changes in the extratropical storm tracks and cyclone intensity, wind speed, and structure. *Weather and Climate Dynamics*, 3, 337–360.

Priestley, M.D.K. & Catto, J.L. (2022b) Improved representation of extratropical cyclone structure in HighResMIP models. *Geophysical Research Letters*, 49, e2021GL096708.

Priestley, M.D.K., Dacre, H.F., Shaffrey, L.C., Hodges, K.I. & Pinto, J.G. (2018) The role of serial European windstorm clustering for extreme seasonal losses as determined from multi-centennial simulations of high-resolution global climate model data. *Natural Hazards and Earth System Sciences*, 18, 2991–3006.

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S. et al. (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168.

Roberts, J.F., Champion, A.J., Dawkins, L.C., Hodges, K.I., Shaffrey, L.C., Stephenson, D.B. et al. (2014) The XWS open access

catalogue of extreme European windstorms from 1979 to 2012. *Natural Hazards and Earth System Sciences*, 14, 2487–2501.

Scaife, A.A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M. et al. (2019) Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, 20, e922.

Scaife, A.A., Kucharski, F., Folland, C.K., Kinter, J., Brönnimann, S., Fereday, D. et al. (2009) The CLIVAR C20C project: selected twentieth century climate events. *Climate Dynamics*, 33, 603–614.

Scaife, A.A. & Smith, D. (2018) A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, 1, 28.

Schwierz, C., Köllner-Heck, P., Zenklusen Mutter, E., Bresch, D.N., Vidale, P.-L., Wild, M. et al. (2010) Modelling European winter wind storm losses in current and future climate. *Climatic Change*, 101, 485–514.

Severino, L.G., Kropf, C.M., Afargan-Gerstman, H., Fairless, C., de Vries, A.J., Domeisen, D.I.V. et al. (2024) Projections and uncertainties of winter windstorm damage in Europe in a changing climate. *Natural Hazards and Earth System Sciences*, 24, 1555–1578.

Stigler, S.M. (1973) Studies in the history of probability and statistics. XXXII: Laplace, Fisher and the discovery of the concept of sufficiency. *Biometrika*, 60, 439–445. Available from: http://www.jstor.org/stable/2334992

Tabari, H. (2020) Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, 10, 13768.

Tyner, B., Aiyyer, A., Blaes, J. & Hawkins, D.R. (2015) An examination of wind decay, sustained wind speed forecasts, and gust factors for recent tropical cyclones in the Mid-Atlantic Region of the United States. *Weather and Forecasting*, 30, 153–176.

Ulbrich, U., Fink, A.H., Klawa, M. & Pinto, J.G. (2001) Three extreme storms over Europe in December 1999. *Weather*, 56, 70–80.

Whetton, P., Macadam, I., Bathols, J. & O'Grady, J. (2007) Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models. *Geophysical Research Letters*, 34, L14701.

Williams, N.C., Scaife, A.A. & Screen, J.A. (2023) Underpredicted ENSO teleconnections in seasonal forecasts. *Geophysical Research Letters*, 50, e2022GL101689.

Woollings, T., Franzke, C., Hodson, D.L.R., Dong, B., Barnes, E.A., Raible, C.C. et al. (2015) Contrasting interannual and multi-decadal NAO variability. *Climate Dynamics*, 45, 539–556.

Zappa, G., Shaffrey, L.C., Hodges, K.I., Sansom, P.G. & Stephenson, D.B. (2013) A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. *Journal of Climate*, 26, 5846–5862.

APPENDIX. RELIABLE SUMMARY OF MODEL TREND ESTIMATE

A reliable single summary of the model trend estimates should account for the different sampling uncertainties of the estimates and also not be unduly influenced by outlier estimates. To achieve this we have summarized the model trend estimates by calculating their weighted median using optimal weights that are proportional to the reciprocal of the standard deviations of the trend estimates provided by the GLM fits.

The weighted median is defined as follows. Denote the model trend estimates sorted into increasing magnitude by $x_1, x_2, \ldots, x_n$ with associated weights $w_1, w_2, \ldots, w_n$, where $w_1 + w_2 + \cdots + w_n = 1$. The weighted median is the midpoint value x_k, where k is chosen such that $w_1 + w_2 + \cdots + w_{k-1} \leq 1/2 \leq w_1 + w_2 + \cdots + w_k$. The weighted median is resistant to changes in the outlier values; for example, it does not change if, say, the smallest value x_1 were even smaller.

The variance of the weighted median is approximated by the reciprocal of $4n(w_1 f_1 + w_2 f_2 + \cdots + w_n f_n)^2$ in the asymptotic limit as $n \to \infty$, where f_i is the probability density of the $i = 1, 2, \ldots, n$th data value evaluated at the median (derived by Laplace in 1818; see Stigler, 1973). The method of Lagrange multipliers with constraint $(w_1 + w_2 + \cdots + w_n)^2 - 1 = 0$ can be used to show that minimum variance is obtained when the weights are proportional to the densities; that is, $w_i = c f_i$.

For Normally distributed data, such as our maximum likelihood trend estimates, densities $f_i = (2\pi)^{-1/2} \sigma_i^{-1}$, and so the optimal weights are $w_i = \sigma_i^{-1} / (\sigma_1^{-1} + \sigma_2^{-1} + \cdots + \sigma_n^{-1})$ (Sec 3.7.1, Brehm, 2023). With this reciprocal standard deviation σ_i weighting, the asymptotic variance of the weighted median is given by

$$\frac{\pi(\sigma_1^{-1} + \sigma_2^{-1} + \cdots + \sigma_n^{-1})^2}{2n(\sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_n^{-2})^2},$$

which is used to calculate the confidence intervals shown in this article.