RESEARCH ARTICLE

Estimating present-day European seasonal mean rainfall by combining historical data and climate model simulations, for risk assessment

Stephen Jewson[1] ⓘ | Francesco Comola[2] | Ben Parkes[3,4]

[1]Lambda Climate Research Ltd, London, UK

[2]Risk Management Solutions Ltd., London, UK

[3]Department of Mechanical, Aerospace and Civil Engineering, University of Manchester, Manchester, UK

[4]Centre for Crisis Studies and Mitigation, University of Manchester, Manchester, UK

Correspondence
Stephen Jewson, Lambda Climate Research Ltd., London, UK.
Email: stephen.jewson@gmail.com

Funding information
Ben Parkes is the recipient of the Ekpe Research Impact Fellowship in the Department of Mechanical, Aerospace and Civil Engineering at the University of Manchester.

Abstract

Building risk models for present-day climate requires an understanding of recent climate trends. To estimate the climate change driven component of recent rainfall trends in Europe, we introduce a novel methodology for combining trend estimates from observed data, a climate model ensemble and a default trend of zero. The methodology weights the different trend estimates based on their uncertainty and consistency with observations. We find that the methodology puts low weights on the observational estimates of recent rainfall trends because they are so uncertain and puts higher weights on the trends estimated using the climate model ensemble mean and the default trend of zero. This demonstrates the value of ensemble simulations of past climate for this application. The methodology we describe establishes a probabilistic framework for estimating uncertain climate change trends based on combining estimates from observed data and climate models and could be applied in many other situations.

KEYWORDS

catastrophe model, climate change, climate model, insurance, model averaging, rainfall trends

1 | INTRODUCTION

Insurance companies are interested in weather and climate predictions for a range of different future time horizons, in order to provide useful information for the different parts of their business. The pricing of insurance generally requires information about the next 0–2 years, reflecting the time periods of most insurance contracts being sold. Planning expansion or contraction of the different parts of the business generally requires information about the next 0–10 years. Choosing how to invest financial reserves may require information about the next

0–30 years. Of these activities, though, it is the first, the pricing of insurance, which is the most pressing, since it most directly influences profitability, and hence the overall viability of the company, in the short term.

How, then, should insurers predict weather and climate in the next 2 years? We use the word predict in a probabilistic sense to refer to estimating the whole distribution of possible outcomes. Following the work of Friedman (1972), the standard approach used in the insurance industry is to combine information from historical weather data and statistical models blended together in appropriate ways to construct large ensemble

simulations. These ensemble simulations consist of tens of thousands of simulated examples of the near-future weather and climate. More recently, this approach has been extended to include information from hydrological, dynamical coastal ocean and dynamical atmosphere models where appropriate (e.g., see the flood model described in Kaczmarska et al., 2018 or the results from Loridan et al., 2013). The resulting simulations of near-future weather and climate are then converted into estimates of possible future physical damage using models for the value and vulnerabilities of buildings. This approach to modelling the impacts of extreme weather on property, for use in insurance risk assessment, is known as catastrophe modelling (cat modelling). Specific examples of catastrophe models (cat models), elements of cat models and applications of cat models have been described in various articles such as Friedman (1972), Vickery et al. (2006), Hall and Jewson (2007), Kaczmarska et al. (2018) and Sassi et al. (2019). An overview of many of the approaches used is given in textbooks such as Grossi and Kunreuther (2005), Mitchel-Wallace et al. (2017) and Michel (2018).

A question that arises in the construction of weather and climate-related cat models is whether climate change should be accounted for in some way. The primary concern is that climate change may have affected the historical weather and climate data that are being used to build the cat models, and if it has, then that data may no longer be relevant for the present and near-future climate without some kind of adjustment. For instance, historical average temperatures from the 1970s are colder than today's average temperatures, and so any risk assessment of temperature that uses the historical data from the 1970s needs to take that into account. A simple assumption that is often applied is that the mean temperature has increased since the 1970s, while the nature of the variability around the mean has not changed (see, e.g., the discussion in chapter 2 in Jewson et al., 2005). The data from the 1970s can then be adjusted by increasing the mean to values estimated to be relevant for the present or near future.

Adjustments based on linear trends, either as a function of time or of global mean surface temperature, are the most commonly used. More complex adjustment schemes, involving non-linear trend adjustments to the mean, and adjustments to the variance, can also be applied. Overall, these adjustments are generally known as detrending. For temperature, detrending is easily justified, since trends in temperature are well understood to be influenced by climate change, and since the size of the trends can be readily quantified. Past sea levels, which also show well understood and readily quantified trends, are similarly straightforward to adjust so that they represent reasonable present or near-future sea level values.

Variables other than temperature and sea level, however, are more challenging to detrend in this way. In this article, we consider rainfall, and the question of whether to, and how to, adjust past rainfall, to account for the possible effects of climate change. The emergence of rainfall trends due to climate change in observations and models varies by region, season and spatial scale (Maraun, 2013). In some regions and seasons, climate change trends in rainfall can be clearly quantified from observations (e.g., Risbey et al., 2013). In others, quantification from observations is much harder because of the weakness of possible trends relative to the 'noise' of climate variability. The challenges of trend quantification in observed rainfall data have been discussed in Jewson et al. (2021). That study compared a number of methods for estimating rainfall trends based on the ideas of statistical testing, model selection and model averaging, and concluded that for insurance risk modelling, trend estimation based on model averaging works better than trend estimation based on statistical testing or model selection.

Jewson et al. (2021) estimated rainfall trends on the basis of observed data alone. However, in addition to trying to identify and estimate trends using observed data, one can also consider climate model simulations as a possible source of information about past rainfall trends. This then leads to three broad categories of methods for estimating trends: data-based approaches, which rely purely on observational data; climate model-based approaches, which rely purely on dynamical climate models; and blended data/model approaches, which combine estimates from both observed data and climate models.

These methods each have pros and cons. Data-based approaches have the advantage that they focus directly on the quantity of interest (actual rainfall) and implicitly capture all the physical effects that control rainfall, including effects that may be missed by climate models. They have the disadvantages that the observed data are very variable so that any long-term trends are often heavily obscured by weather and climate variability, and that in some regions, observations may have very limited spatial resolution or short historical records. In regions where there are no observations at all, data-based methods clearly do not work.

Climate model-based approaches have the advantages that climate models can be run as ensembles, and that models provide information in regions where there are few observations. The benefit of ensembles arises if the separate ensemble members are initialized in such a way that the weather and climate variability in the different ensemble members is randomly out of phase. In this case,

JEWSON ET AL.

Meteorological Applications
Science and Technology for Weather and Climate
Open Access

RMetS

3 of 17

when the ensemble members are averaged together to create the ensemble mean, the noise due to weather and climate variability reduces, while any underlying trends remain the same. The signal-to-noise ratio for estimating trends is thus improved, and estimates of trends from climate model ensembles are therefore typically more precise (although not necessarily more accurate) than estimates of trends from observational data. The disadvantage of climate model-based methods is that all aspects of climate model simulations are only an approximate representation of reality. In particular, real precipitation is contingent on processes that are smaller scale than climate model resolution, and as a result precipitation in climate models has to be parameterized. Any estimates of trends from climate models should therefore be considered to be possibly biased (see, e.g., discussions about the sensitivity of climate model results to rainfall parameterizations in Knutson & Tuleya, 2004, Birch et al., 2014, Garcia-Carreras et al., 2015, Willetts et al., 2017 and Konduru & Takahashi, 2020, and a discussion of biases in climate model precipitation in Maraun, 2012).

In this article, our goal is to derive estimates of unknown climate change trends by drawing on information from both observations and climate models. We will test a novel methodology for combining estimates from observations and climate models that attempts to produce trend estimates that are better than can be produced from either alone. The method we test works by considering various factors: the observed trend, the uncertainty around the observed trend, the climate model ensemble mean trend, and the consistency between the climate model trend and the observed data, taking into account the level of variability in the data. The analysis of these factors and the combination of models is done in a single statistical calculation based on standard probabilistic principles.

Our goals are similar to, but also different from, the usual goals of the well-established practice of downscaling of climate models. Downscaling is concerned with using observations to adjust the outputs of climate models to make them more realistic (see, e.g., Maraun & Widmann, 2018). Cat models are very often not, however, created directly from climate model output, but rather from statistical simulations of weather and climate (in order to be able to simulate many more years of realistic variability than is possible with climate models). Our goal is to understand how best to derive estimates of the climate change trend from observations and climate models that can then be used to adjust these statistical simulations.

This study is part of a wider effort to create principled mathematical and statistical methods for extracting quantitative information from climate model output and

observed data that can then be used to improve the modelling of a number of the variables that are used in cat models, including various aspects of rainfall, but also including, for example, wind speeds and hurricane numbers. We believe that there is great potential for this approach to lead to better risk models. Better risk models will help insurers price and manage risk better, leading to a more efficient insurance industry. A more efficient insurance industry can help society as a whole to manage extreme weather risk more effectively, especially as the climate changes.

In Section 2, we describe the data and climate models that we will use. In Section 3, we discuss the question of how observational data and output from climate models can be combined, and introduce the model averaging method that we will test. In Section 4, we present results for a single rainfall index: Spanish summer mean rainfall. In Section 5, we present results for other parts of Europe. In Section 6, we explore the sensitivity of the results to some of the methodological assumptions, and in Section 7, we summarize and conclude.

2 | OBSERVED CLIMATE DATA AND CLIMATE MODEL SIMULATIONS

We will use both observed rainfall and climate model simulated rainfall to create trend estimates, which we then combine in a logical way. For both the observed and simulated rainfall, we aggregate the data in space and time to create summer and winter mean rainfall indices for eight large geographical areas in western Europe, namely: UK (UK), Northern France (FN), Southern France (FS), Spain (ES), Italy (IT), Northern Germany (DN), Southern Germany (DS) and Scandinavia (Denmark, Sweden and Norway) (DK). We will denote summer and winter indices as, for example, UKs and UKw. Our 16 regions follow the regions used in Jewson et al. (2021).

Ultimately, we are interested in deriving best estimates of trends in rainfall at a greater level of granularity than just seasonal mean indices on large scales and for extreme rainfall as well as mean rainfall. However, in this article, we will focus on seasonal mean rainfall as a starting point for the development of concepts and the exploration of methodologies for combining estimates from climate models and observations.

2.1 | Observed data

The observed rainfall data we use are from the E-OBS dataset (Cornes et al., 2018). These data consist of rainfall

station observations interpolated onto a 25 km grid. Data are available from 1950 to 2018, but in this study, we only use data from 1981 onwards to match the climate model output (which starts in 1981), giving 38 years in total.

2.2 | Climate model data

The climate model output we use was produced by the United Kingdom Meteorological Office (UKMO) as part of the 2018 iteration of the United Kingdom Climate Projections (UKCP18) project (Lowe et al., 2018). We use an ensemble of 12 limited-area atmospheric model runs at 12 km horizontal resolution, produced by the UKMO Unified Model Global Atmosphere GA7, and driven by perturbed variants of the UKMO global climate model HadGEM3-GC3.05. These runs cover Europe for the period 1981–2080. We will use data from the period 1981–2018 to match the observations (which end in 2018). The boundary conditions for the climate model simulations were observed forcings including, in particular, increasing greenhouse gases. The climate model runs were started from different initial conditions in 1900, and use random numbers to perturb the physical parameterizations (see the explanation in Lowe et al., 2018). As a result of the initialization, the model states of the different ensemble members during the period 1981–2018

would be expected to be randomly out of phase with respect to each other in terms of decadal and interdecadal climate variability, as we require for our study.

These climate model simulations use just a single global model and a single regional model to produce the ensemble. This type of ensemble is known as a single model ensemble (SME). An alternative approach is to use a selection of different global models and/or a selection of different regional models to produce the ensemble, which is known as a multi-model ensemble (MME). SMEs and MMEs have various advantages and disadvantages. MMEs have the advantage that they capture more of the model uncertainty, by not relying on individual models, and hence the ensemble spread may be larger. However, they have the disadvantage that the question of what weights to put on each model has to be considered, and although various methods have been suggested for how to derive such weights, no general consensus has been reached. In our application, the main algorithm we study (see Section 3.5) only uses the ensemble mean, and hence even if the SME we use does underestimate the ensemble spread, that would have no impact on the results. All of the methods we describe below could be applied equally well to SMEs and MMEs.

Time series for Spanish summer mean rainfall for four ensemble members are shown in Figure 1, along with straight line trends fitted using ordinary least
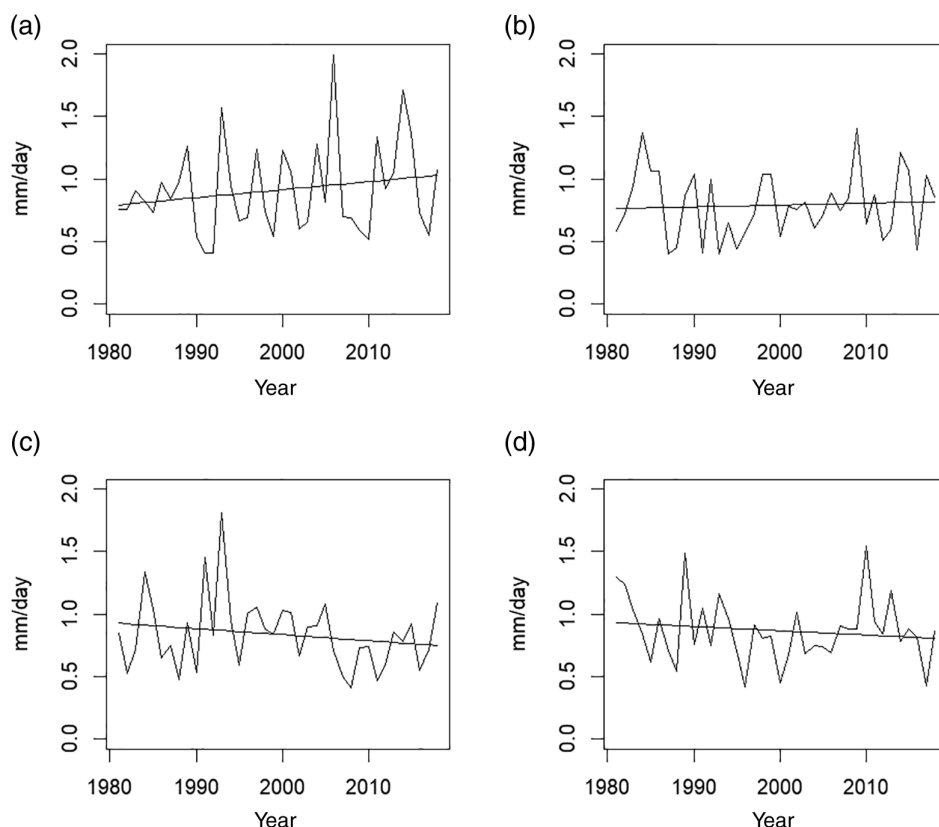

FIGURE 1 Spanish summer mean rainfall for four ensemble members from the climate model, with ordinary least squares (OLS) trend
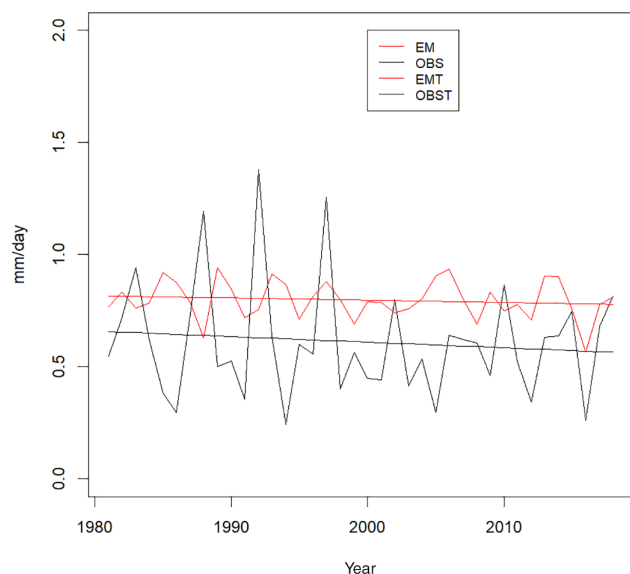
FIGURE 2 Spanish summer mean rainfall for the climate model ensemble mean (EM), and the observations (OBS), along with OLS trends fitted to both (EMT and OBST, respectively)

squares (OLS). We see that the variability is indeed different (i.e., out of phase) in the different ensemble members. The OLS trends are also different: some are increasing, and some are decreasing. These trends are due to a combination of any climate change trend, which would be the same in each ensemble member, plus any residual trends due to interannual or decadal climate variability, which would be different in each ensemble member.

Figure 2 shows the time series of the ensemble mean rainfall (in red) and the time series of the observed rainfall for the same period (in black), also for Spanish summer mean rainfall. Figure 2 shows that the model ensemble mean shows a clear bias in the mean level of rainfall relative to the observations, with the model around 30% higher. Across our 16 indices, this bias varies from region to region and winter to summer. Figure 2 also shows OLS trends fitted to the ensemble mean and observed time series. Both show decreasing trends. That they are both decreasing may be a coincidence, but decreasing trends in Spanish summer rainfall are also seen in the projections for future changes in rainfall given by EURO-CORDEX (Jacob et al., 2014). It is tempting to ask whether the trends in these time series are statistically significant. However, consideration of statistical significance is not particularly relevant when the goals are to create a best estimate of the trend, accurate and low volatility predictions, and inputs for risk models. This is discussed in detail in Jewson et al. (2021) and in Section 3.1. Instead of focusing on statistical significance, we will apply a model averaging methodology that considers the size of the trend signals, their uncertainty and the

consistency of the trends with the observed data to create a best estimate of the unknown climate change trend.

We conclude from the biases shown in Figure 2 that the mean levels of rainfall from the model should be discarded, as is usually the case in downscaling methods, and we will determine the mean level of rainfall purely from the observations. We note, however, that the biases between model and observations may not only be due to deficiencies in the climate model. The observations themselves also have limitations, mainly due to the finite network of stations on which rainfall observations are made (Herrera et al., 2019).

One of the challenges involved in interpreting climate model output is understanding whether or not the biases in the mean should be taken as indication that the model is sufficiently unrealistic that nothing from the climate simulations can be trusted, or whether it is possible that even though the mean climate is incorrect, some aspects of climate variability may still be modelled correctly. In our case, the question is whether the rainfall trend response to changing greenhouse gases can be considered correct if the mean rainfall is not correct. The statistical methods we will use for combining observed rainfall data with climate model rainfall will address this issue, by taking account of the extent to which the climate model trend can be considered consistent with observed data. To the extent the climate model trend is not consistent with observed data, the trend will be downweighted. To the extent it is consistent, then it will be combined with the observed trend, and the possibility of no trend, each with an appropriate weight. This occurs automatically in the statistical methods that we will use.

3 | COMBINING CLIMATE MODEL OUTPUT WITH OBSERVATIONS

There are many uses that one can make of the observational and climate model data described above. The goal of this particular study is to try and derive a reasonable estimate of the trend in the seasonal mean rainfall over the recent period, by combining information from observed and climate model datasets in a way that puts appropriate weights on the two sources of information. How one might go about combining the two depends on one's assumptions about climate models and their ability to simulate observed variability and trends. In our approach, we assume that the climate model may contain useful information but that it also has to 'prove' itself by being consistent with observations before it is used. We now discuss in more detail why we choose not to use statistical significance testing, and then describe the model combination methodology that we will apply.

3.1 | Using statistical testing

We will see below that the trends in our rainfall data are weak and mostly not statistically significant. We could therefore, perhaps simply model the trends as zero until they become significant at some point in the future, if they ever do. We will not use this approach, however, for the following two reasons.

First, statistical testing is designed as a binary test that gives low numbers of false positives (a.k.a. type I errors) and high numbers of false negatives (a.k.a. type II errors). When using statistical testing on trends, the risk of accepting a spurious trend is low, while the risk of rejecting a real trend is high. This is an appropriate balance for the purposes of scientific discovery. We, however, have a different goal, which is simply to make an accurate-as-possible estimate of the trend, to enable accurate estimates of current and near-future climate. Since this is a different goal to scientific discovery, it is not surprising that different tools are required. Jewson et al. (2021) have analysed the pros and cons of statistical testing and compared statistical testing with other methods for estimating trends and have shown that model averaging methods outperform statistical testing in terms of predictive accuracy when modelling weak trends.

Second, even if the trends are not significant now, they may become significant in the future (see, e.g., the discussion of emerging rainfall trends in Maraun, 2013). If one is using statistical testing, then at the point in time at which a trend becomes significant, the estimate of the trend changes somewhat dramatically (see the example in Jewson et al., 2021), and this will occur at different times for different seasons and indices. These dramatic changes are unfortunate for risk modelling, and it is preferable to establish a methodology from the start that will detect emerging trends at an early stage, even at the risk of modelling spurious trends. This argument is further justified by the simulation results in Jewson et al. (2021), which show that model averaging methods give less volatile predictions than statistical testing.

3.2 | AIC model averaging

We will now describe in detail what we believe is a new method for assessing and combining estimates of trends from models and observations to create a best estimate. The method is based on the information theory developed by Akaike, as discussed in atmospheric science textbooks such as von Storch and Zwiers (1999) and Wilks (2011), and statistics textbooks such as Wasserman (2004), Burnham and Anderson (2002), Claeskens and Hjort (2008) or Fletcher (2019). Akaike's theory gives a score for

estimating how well a statistical model fits the data to which it is fitted, where the score includes a penalization factor proportional to the number of parameters in the model, to avoid overfitting. This then gives a way of measuring how close the fitted model is to the unknown truth. This theory is most commonly used for selecting one from a number of statistical models (i.e., model selection). However, it can also be used as a simple but principled way to combine models (i.e., model averaging).

Akaike's theory can be applied when a number of parametric statistical models have been used to model the same dataset. In our case, the dataset being modelled is the observation time series for one of our 16 indices, and the parametric statistical models are flat-line (i.e., no trend), a linear trend fitted to the data, and a model based on the linear trend from the climate model ensemble mean. These models are described in more detail below. To apply Akaike's theory for assessing and combining models, each model will be fitted using maximum likelihood. The likelihood is a standard measure of goodness of fit between a model and data, with higher values indicating a better fit, and the log-likelihood achieved at the maximum by each model can be used to measure how well that model has fitted the data. However, the log-likelihood achieved at the maximum cannot be used to compare the appropriateness of different models when the models have different numbers of parameters, since models with a greater number of parameters will naturally tend to fit the data better and hence achieve a higher value of the log-likelihood. This affects us because our three models do have different numbers of parameters: the flat-line and linear trend models have two and three fitted parameters, while the ensemble mean trend model has two parameters. We could not, therefore, use the log-likelihood on its own to determine which of the three models would be likely to give the best predictions. Akaike derived a correction for the overfitting effect of having extra parameters and introduced a simple score, which is based on the log-likelihood but includes a penalization proportional to the number of parameters. This score can then be used to compare models, such as our three models, even though they contain different numbers of parameters. The original score introduced by Akaike is known as the Akaike information criterion (AIC) and is given by:

$$\text{AIC} = 2k - 2\ln L, \qquad (1)$$

where AIC is the AIC score, L is the likelihood achieved at the maximum, and k is the number of free parameters that were adjusted to maximize the likelihood in that model. The sign convention is such that *lower* values indicate better models (which is the opposite sign convention to likelihood).

In our case, because we are using linear, univariate models with normal distributions, it is appropriate to use a version of the score known as the AICC, which includes a small-sample correction, which is given by:

$$\text{AICC} = \text{AIC} + \frac{2k^2 + k}{n - k - 1}, \qquad (2)$$

where n is the number of data points. The AIC and AICC scores are most commonly used for selecting the most appropriate model from a set of models. However, they can also be converted into weights and used to create a new model based on the weighted average of the individual models, with the intention that the combined model is better than any of the individual models. Given the AICC scores for each model, the weights for the individual models that are used to create the combined model are calculated using the following steps:

a. Define the minimum of the AICC scores across the models as AICCmin.
b. Calculate the deviations of each of the model scores from this minimum, DAIC = AIC − AICCmin.
c. Calculate the *relative likelihood* of each of the models as RL = $\exp\left(-\frac{\text{DAIC}}{2}\right)$.
d. Normalize the relative likelihoods so that they sum to one, to create weights.

Further details of AIC and AICC scoring and weighting are given in the textbooks by Burnham and Anderson (2002), Claeskens and Hjort (2008) and Fletcher (2019).

When we put weights on our three models using this method, we will not assume that the trends in different regions and seasons are related, since both observational and model studies show rainfall trends that vary significantly by region and season (see, e.g., the results shown in Jacob et al., 2014). We will also not assume that the weight to be applied to the climate model is the same for different regions and seasons, and we will calculate it separately for each index. This is appropriate since the rainfall climate in the different regions and seasons is governed by different physical processes (such as different types of clouds, different temperature climates and so on) and the climate model may simulate some well and others less well.

We now describe in more detail the three models that we will blend together using AIC theory.

3.3 | Flat-Line model (FL)

The first model in the set of models that we will combine together to create a best estimate of the trend in mean rainfall is a model that we call flat-line (FL), which consists of simply modelling the rainfall time series as stationary. We will assume that the deviations around the mean are normally distributed, as well as independent and identically distributed, and so this model corresponds to nothing more than fitting a normal distribution to the data, fitted using maximum likelihood. We discuss using alternative distributions in Section 7. This model uses two parameters: the mean and the standard deviation. We can write this model as follows:

$$r_t = \mu + \sigma \epsilon_t, \qquad (3)$$

where r_t is the mean seasonal rainfall in year t; μ is the mean over all the years of the mean seasonal rainfall values; σ is the standard deviation of deviations around the mean, and ϵ_t is a standard normal random variable that represents the individual standardized fluctuations around the mean in year t. The two parameters to be estimated from the observed rainfall data are μ and σ. We include this model because for the time series we are considering the trends are possibly sufficiently weak, relative to the variability, that it may be better not to model them at all.

Data do not have to be precisely stationary for this to be the best model. They just have to be sufficiently close to stationary that there is insufficient information to model the non-stationary aspects well enough to improve predictions.

3.4 | Observed trend model (OBST)

The second model in the set of models that we will combine together is a model that consists of fitting a linear trend to the observed rainfall time series, again using maximum likelihood. This model uses three parameters: the mean, the trend slope and the standard deviation.

In this case, fitting maximum likelihood is equivalent to fitting the mean and slope using OLS and then calculating the standard deviation of the residuals around the trend. We can write this model as follows:

$$r_t = \mu + \beta (t - \bar{t}) + \sigma \epsilon_t, \qquad (4)$$

where β is the trend slope. The three parameters to be estimated from data are μ, β and σ.

It is not clear a priori which of the FL and OBST models will be deemed closer to the truth by the AICC score. The OBST model will give a higher (i.e., better) likelihood value because it has more parameters, and hence fits the data more closely, but in the AICC score, it will be penalized for the extra parameter. Which model

has the lower (i.e., better) AICC score will therefore depend on the size of the trend relative to the variability. If the trend is weak, then FL may achieve a lower AICC score, indicating that better predictions would likely be made by ignoring the trend rather than trying to estimate it.

One could also consider non-linear models for the trend in the rainfall data, since the trend is undoubtedly not completely linear. However, even in the OBST model, the slope parameter is hard to estimate because of the low signal-to-noise ratio, and additional parameters would be even harder to estimate.

One could also model the trends by modelling the rainfall as a function of global mean surface temperature (GMST) rather than time. This would be necessary if longer time series of historical rainfall were being used, but since GMST evolution is close to linear as a function of time over the time period that we are considering, it would be unlikely to make much difference in this case.

3.5 | Ensemble mean trend model (EMT)

The third model we include in our model set uses the climate model output to model the observed rainfall data. As discussed above, the mean of the climate model output is biased relative to the observations, and so we discard it. We also discard the climate model variability, since all we are interested in from the climate model output in this study is the estimate it provides of the trend. The individual ensemble members each have their own trend. We saw in Figure 1 that these trends vary significantly because they are strongly influenced by variability within that particular model run. By averaging the members together to create the ensemble mean, much of this variability cancels out, and the ensemble mean series therefore gives a better estimate of the climate change trend in the rainfall in the model. In the EMT model, we then model the observed data using the mean estimated from the observed data, the trend estimated from the climate model ensemble mean, and the standard deviation estimated from the observed data. The only difference between this and the OBST model above is that the trend estimate comes from the climate model. This model can be written as follows:

$$r_t = \mu + \beta_{CM}(t - \bar{t}) + \sigma \epsilon_t, \qquad (5)$$

where β_{CM} is the trend from the climate model ensemble mean. The parameters to be estimated from the observed rainfall data are μ and σ, but not β_{CM} since that is given by the climate model ensemble mean. Crucially for our

model averaging method to work, the EMT model only counts as having two free parameters for the AICC score. This is because the trend is not estimated using the observed data we are trying to model, but comes from a separate external dataset (the climate model output). From the point of view of maximizing the likelihood of the observed data, the climate model trend is a given constant, not an estimated parameter.

We note that we do not scale the climate model trend to get it to match the observed trend by applying a bias correction to the climate model trend. There would be no point in doing this: the resulting model would be equivalent to the OBST model, and the climate model data would no longer play any role. If there is a bias in the climate model trend, that will be reflected in the weight applied to the EMT model.

The AICC score for the EMT method may be higher (worse) or lower (better) than that for the OBST model. On the one hand, since EMT only uses two parameters rather than three, we might expect it to have a lower AICC score, and hence beat the OBST model. On the other hand, the trend used by the EMT model will not fit the observed data as well as the trend used by the OBST model, since the trend used by the OBST model is by definition the best possible fit to the data. As a result of this effect, the EMT method will tend to score a lower likelihood than the OBST method, which will increase the AICC score. Which of the OBST and EMT models has the lower AICC score in the end, and hence which of the models would be expected to give better predictions, then comes down to a balance between these two effects.

3.6 | Weighted combinations

We can combine the three models described above in four different ways, consisting of three two-way combinations and one three-way combination. The first combination we consider is the two-way combination between FL and OBST and ignores the climate model. This combination considers how well the trend can be estimated in the observational data and, taking into account the uncertainty around the estimate, combines the two models accordingly. If the trend can be well estimated in the observational data, then OBST will get most of the weight in this combination. If the trend is poorly estimated, then FL will get most of the weight. The shortcoming of this method is that it makes no attempt to incorporate output from the climate model as a possible source of information.

The second model combination we consider is the two-way combination between OBST and EMT. This combination considers which of the two methods for

estimating the climate change trend is the better one: fitting the trend to the data or taking the trend from the climate model. If the trend estimate from the climate model is not consistent with the observed data, even when taking into account the high level of noise in the data, then OBST will get most of the weight. On the other hand, if the trend estimate from the climate model is consistent with the observed data, then EMT will get most of the weight because OBST is penalized for having an extra parameter. The shortcoming of this method is that it fails to account for the possibility that neither trend model is very good. It does not allow for the possibility that it might be better to use a trend closer to zero than the trend from either of these models.

The third combination we consider is the two-way combination between FL and EMT. This combination could be useful in situations in which the observed trend is poorly estimated, because of the level of noise in the data, and hence is perhaps better ignored. It weights both FL and EMT according to their consistency with the data. If EMT captures the correct sign for the observed trend, it will beat FL.

The fourth, and most interesting, combination that we consider is the three-way combination among FL, OBST and EMT. This combination takes into account all the factors discussed above and weights the three models
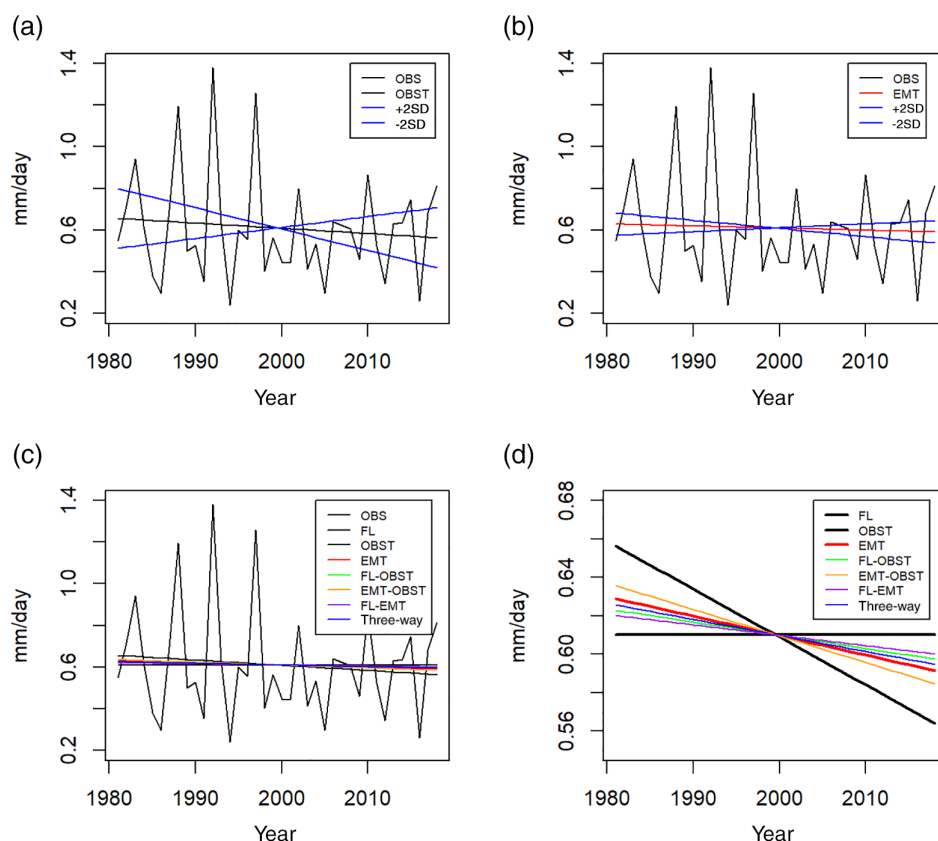
appropriately. It chooses automatically, and on a reasonable basis, whether to model a trend or not, and if the trend is modelled, how much to follow the observed trend and how much to follow the EMT. Based on the underlying statistical theory, one would expect that this three-way combination would likely give better predictions than any of the individual models or any of the two-way combinations.

4 | SPANISH SUMMER RAINFALL RESULTS

We now present results from applying the methods described above to Spanish summer rainfall. Figure 3a,b shows the uncertainties around the trend estimates shown in Figure 2, using trend lines with slopes that show 95% confidence intervals above and below the observed and climate model trends. Panel (a) shows the uncertainty around the observed trend: the uncertainty is large, and the trend is not statistically significant (although statistical significance does not affect our model combination analysis, as discussed above). This large uncertainty is an indication that the weight given to the observed trend in our multi-model combinations may not be very high. Panel (b) shows the uncertainty around

FIGURE 3 Panel (a): Observations of Spanish summer mean rainfall (OBS), with fitted OLS trend (OBST) and with two additional lines (+2SD, −2SD) with the same mean as the OLS trend but with the trend increased and decreased by two standard errors, to illustrate the uncertainty on the trend. Panel (b): As panel (a), but for the trend from the climate model ensemble mean (EMT). Panel (c): Observations of Spanish summer mean rainfall (OBS), along with seven methods for estimating the trend. The individual methods are explained in the text. Panel (d): the seven trends as shown in panel (c) but now with an expanded vertical scale. The three thicker lines correspond to the fundamental methods (FL, OBST and EMT) and the four thinner lines correspond to the four model combinations

the EMT. The line representing the EMT has been shifted downward so that the mean matches that of the data, reflecting that the EMT model uses the mean of the observations. We see that the uncertainty on the EMT is smaller than the uncertainty on the observed trend: this is the benefit of having an ensemble average. The EMT is also not statistically significant.

Table 1 gives various statistics for the three models, applied to the Spanish summer rainfall time series. The first row shows the sizes of the estimated trends. The OBST trend is more than twice the EMT trend. The next four rows show terms in the expression for the AICC. The second row shows the log-likelihoods. We have multiplied the log-likelihoods by minus two, since they are multiplied by minus two in the expression for the AICC (Equation 2). This means that the numbers in the table are negatively oriented, with lower values indicating a better fit to the data. We see that OBST has the lowest (best) value. This is to be expected, for all 16 cases, since OBST has more parameters that can be adjusted to fit the data. EMT has the second lowest value. This is not necessarily to be expected, but occurs in this case because the climate model trend has the same sign as the observed trend, and hence EMT improves the fit to the data relative to the FL model, which does not use a trend. The third row in the table gives the numbers of parameters in the models, multiplied by two since the number of parameters is multiplied by two in the expression for the AICC. The fourth row gives the small-sample correction that converts AIC to AICC (Equation 2). The fifth row gives the AICC values, which are the sum of the second, third and fourth rows. These are the values that can tell us which model is likely to give the best predictions. We see that EMT has the lowest AICC value, and is hence chosen as the best model. FL is a close second, and OBST

is a distant third. From the contributions of the different terms to the AICC value, we can see that the poor performance of OBST is driven mainly by the penalty for adding a third parameter, which greatly outweighs the differences in likelihoods. The observed trend is being deemed by the combination method as unreliable as an estimate of the climate change trend, relative to the other methods. If we were using a model selection philosophy, rather than a model averaging philosophy, and hence aiming to choose a single 'best' model, then at this point we would choose EMT since it has the lowest AICC value. The sixth and seventh rows of Table 1 relate to sensitivity tests discussed in Section 6.

Table 2 gives the weights assigned to the three models when they are combined in the four different possible combinations using the AICC weighting methodology. In the FL-OBST two-way combination, FL gets 72% of the weight, and OBST only 28%. This is because FL has a distinctly lower AICC value, which is mainly caused by FL having one fewer parameter than OBST. That the OBST model captures the trend using its extra parameter does not increase its AICC score much because the trend is weak relative to the variability in the data, and so capturing the trend correctly does not explain much more of the variability in the data than not capturing it. In the OBST-EMT two-way combination, EMT achieves 75% of the weight, to 25% for the OBST model. This is because EMT has a distinctly lower AICC value than OBST, driven by the fact that it uses only two parameters relative to the three parameters used by OBST, and yet explains the data nearly as well. In the FL-EMT two-way combination, the two models get roughly equal weights: FL gets 47% of the weight, and EMT gets 53% of the weight. These two models have the same number of parameters. EMT gets a slightly higher weight because it has a slightly lower AICC value, driven by the fact that the EMT model captures the correct sign of the observed

TABLE 1 For Spanish summer mean rainfall, for the flat-line (FL), observed trend (OBST), ensemble mean trend (EMT) and ensemble mean trend relative trend (EMT-RT) models: The estimated trend, minus two times the log-likelihood, two times the number of fitted parameters, the small-sample correction, the AICC score, the BIC score and the WAIC score

	FL	**OBST**	**EMT**	**EMT-RT**
Trend	0.000	−0.095	−0.038	−0.029
−2 log-likelihood	4.452	4.021	4.173	4.226
2*# parameters	4	6	4	4
Small sample	0.343	0.706	0.343	0.343
AICC	8.795	10.727	8.516	8.569
BIC	11.728	14.933	11.449	n/a
WAIC	9.521	10.821	9.238	n/a

Note: The AICC score in row 5 is the sum of rows 2, 3 and 4.

TABLE 2 For the FL, OBST and EMT models, the weights in eight model combinations

	FL	**OBST**	**EMT**	**Trend**
FL-OBST	0.724	0.276	0.000	−0.026
OBST-EMT	0.000	0.249	0.751	−0.052
FL-EMT	0.465	0.000	0.535	−0.021
Three-way basic	0.395	0.150	0.454	−0.032
Three-way RT	0.400	0.152	0.448	−0.028
Three-way BIC	0.425	0.086	0.489	−0.027
Three-way WAIC	0.374	0.195	0.431	−0.035
Three-way prior	0.566	0.108	0.326	−0.023

Note: In the final column, the trend for each of the model combinations.

JEWSON ET AL.

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

11 of 17

trend, while the FL model has no trend at all and so is slightly less consistent with the observed data. In the three-way combination FL-OBST-EMT, EMT and FL get the highest weights, of 45% and 40%, while OBST gets only 15%. This can be explained using the reasons given above for the performance of the different models in the two-way combinations.

Figure 3c,d shows the trends estimated from the four combinations, which could be used to predict near-future values. The different trends are hard to distinguish in panel (c) and so are shown with an expanded vertical axis in panel (d). The sizes of the trends are also given in Table 2. The bold lines in Figure 3d show the three fundamental trend models: FL, OBST and EMT. The fine lines show the weighted combinations. The weighted combinations necessarily lie within the range of the three fundamental trend models. Three of four of the weighted combinations lie between EMT and FL. The only combination with a larger trend than EMT is the combination that does not include FL.

5 | EUROPE-WIDE RESULTS

We now present results for our 16 indices, covering 8 European regions, for winter and summer. Figure 4 shows the observed trends and the EMTs for all 16 indices, along with uncertainties given by 95% confidence intervals. Both observed and modelled trends are a mix of positive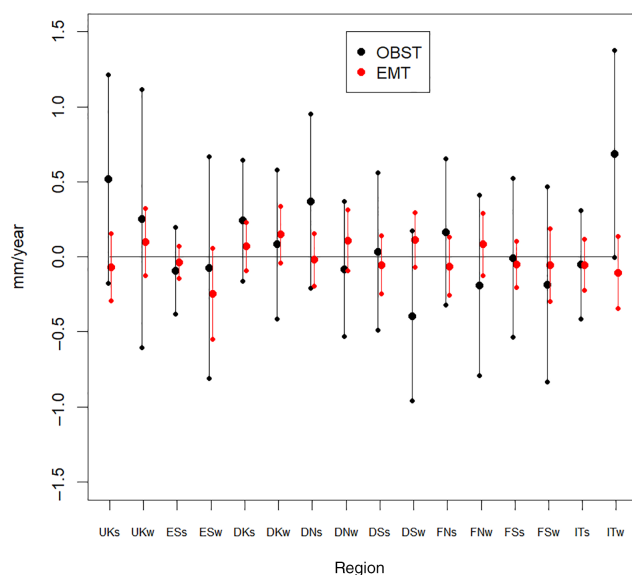 and negative values. All the values show significant uncertainty, and none are significantly different from zero (although the observed trend for Italy in winter is very close). The uncertainties around the climate model EMTs are much less than those around the observed trends because of the averaging involved in calculating the ensemble mean. There is not much correspondence between the trends from observations and the climate model: in fact, for eight cases, the trends show the same sign, and for eight cases, they show different signs.

We can compare the signs of these trends with other studies. Comparing with EURO-CORDEX projections for the long-term future impact of climate change on European seasonal rainfall (Jacob et al. 2014), we see that the sign of the UKCP climate model agrees in almost all cases. The only cases where there is no clear agreement between UKCP and EURO-CORDEX are for (a) ESw, where UKCP suggests a positive trend, while EURO-CORDEX is ambiguous because the north and south of Spain show different sign trends, (b) ITw, where UKCP suggests a positive trend but again EURO-CORDEX is ambiguous, and (c) FSw, where UKCP suggests a negative trend, but EURO-CORDEX shows a positive trend. This comparison suggests that the lack of agreement between the trends estimated from the observations and the climate model is more likely to be due to uncertainty in the observed trends than it is due to biases in the climate model.

5.1 | Combining FL and OBST

Figure 5a shows the weights that result from applying the AICC model weighting methodology to combine the FL and OBST models in a two-way combination, separately for all 16 locations. We have already discussed the results for ESs in Section 4. In all cases except Italy winter, FL gets larger weights. The small weights on the OBST model reflect the small size of the observed trends, relative to the size of the variability, resulting in a high level of uncertainty on the trend estimate. The weights on OBST are in a range from 0.2 to 0.5. This means that the model combination method is recommending that we use the observed trend, but divided by a reduction factor that varies from around 5 to around 2. For Italy winter, OBST gets larger weights. This reflects the large size of the observed trend relative to the uncertainty in this case.

5.2 | Combining OBST and EMT

Figure 5b shows the weights that result from applying the AICC model weighting methodology to combine the

FIGURE 4 Observed and climate model ensemble mean trends for 16 indices from 8 countries and 2 seasons, all with uncertainty of plus and minus two standard errors illustrated by the lines and smaller dots
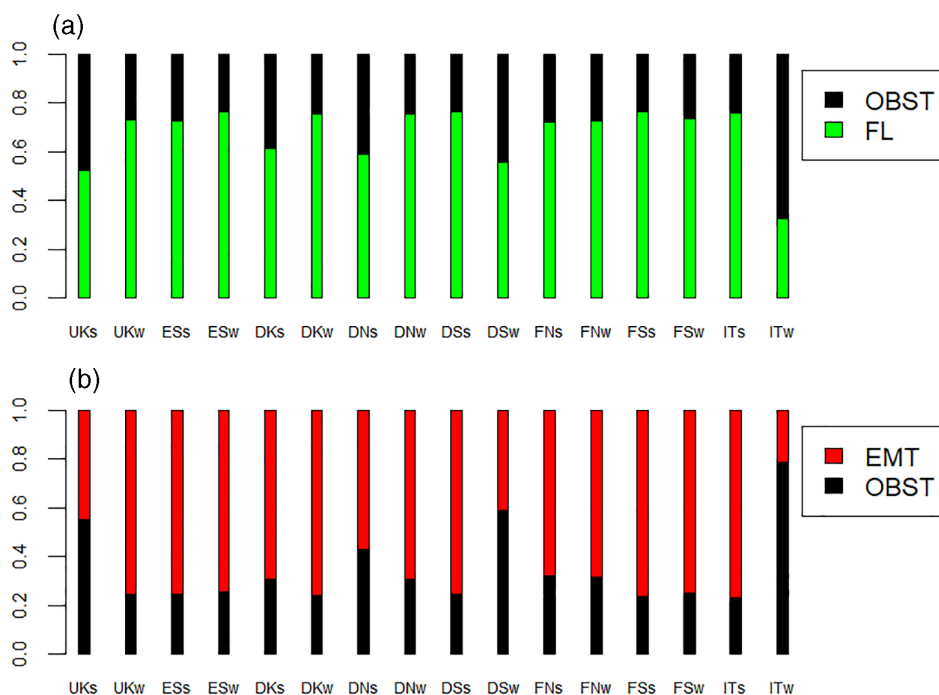
FIGURE 5 Weights from model averaging for the 16 indices. Panel (a) shows the weights for the combination of the FL and OBST models, and panel (b) shows the weights for the combination of the EMT and OBST models

FIGURE 6 Weights from model averaging for the 16 indices. Panel (a) shows the weights for the combination of the FL and EMT models, and panel (b) shows the weights for the three-way combination of the FL, EMT and OBST models

OBST and EMT models in a two-way combination. In 13 of 16 cases, EMT gets larger weights than OBST. The low weights on OBST once again reflect the small size of the observed trend relative to the variability. The three cases where OBST gets larger weights than EMT (UKs, DSw and ITw) are all cases where the size of the observed trend is relatively larger and is of the opposite sign to the climate model trend.

5.3 | Combining FL and EMT

Figure 6a shows the weights that result from applying the AICC model weighting methodology to combine the FL and EMT models in a two-way combination. Since both models have two parameters, the combination is based purely on the likelihood, which measures the consistency of the two models with the observed data.

Overall, the models are given roughly equal weights. Those cases in which EMT gets relatively higher weights (e.g., DKs, FSw, ITs) correspond to cases where the sign of the climate model trend agrees with the observed trend. However, it is perhaps surprising the EMT still gets weights nearly as large as FL even in cases where the climate model and observed trends disagree (e.g., DNw, DSw). This is because the evaluation of the climate model trend is not based on a comparison between the climate model trend and the observed trend, but on the EMT model as a whole, and its ability to represent the entire distribution of observed values. For instance, in the case of DSw, even though the climate model trend is of the opposite sign to the observed trend, the climate model trend is weak relative to the variability, and the EMT model does capture the distribution of variability. As a result, the EMT model is not too inconsistent with the observations, even with the wrong sign of trend. The AICC method states the following: although the climate model trend is of the opposite sign to the trend in the observations, we cannot say that it is incorrect, because of the large variability in the data. Based on the statistical evidence, from comparing the model with observations, the climate model trend is still a plausible candidate for what the trend might be and should be given some weight accordingly. This illustrates how the AICC combination method considers consistency between model and observations in an appropriate probabilistic sense, incorporating uncertainty.

5.4 | Combining FL, OBST and EMT

Figure 6b shows the results of using the AICC model weighting methodology to combine the FL, OBST and EMT models in a three-way combination. The weights in the three-way combination arise from an interplay between a number of factors including the size and uncertainty of the observed trend and the extent to which the EMT model is consistent with the observations. Overall, the weights on the observed trend are low: in 12 of 16 cases, the weights are below 0.2, meaning that in the combined prediction, the observed trend slope is divided by 5 or more. This is how the model averaging method deals with statistically insignificant trends: rather than setting them to zero, it assesses the level of information about the trend and reduces the trend estimate accordingly. In only one case (ITw) does the observed trend capture a weight of more than 0.5. In all 16 cases, the FL and EMT models capture roughly equal weight. As in the FL versus EMT comparison discussed in Section 5.3, EMT tends to get more weight in the cases where the climate model trend agrees with the observed trend (such as

DKs), and less when they tend to disagree (such as DSw). Again, EMT never gets particularly small weights, even when the trend in the climate model disagrees with the trend in the observations, because even with the wrong sign trend, the EMT model as a whole is still reasonably consistent with the observations.

5.5 | Trend estimates

Trend estimates from all the model combination methods are shown in Figure 7, along with the trends from the three fundamental models (FL, OBST and EMT). The trend estimates from the combination methods are bounded by the three fundamental models, as we would expect.

We will discuss the trend estimates from the three-way combination in detail. In eight cases of 16, the OBST and EMT trends agree in sign, and in all these eight cases, the three-way combination has the same sign. In three of these eight cases, the three-way combination lies in-between OBST and EMT, while in five of these eight cases, it lies outside and closer to the FL model (i.e., closer to zero). These five cases are cases in which significant weight has been given to the FL model, in preference to the OBST model, because the signal-to-noise in the observed trend is low. In eight cases of 16, OBST and EMT trends disagree in sign. In all these cases, the three-way trend lies between the OBST and EMT estimates, as it must. It lies on either side of the FL model, and the trend is generally rather small.
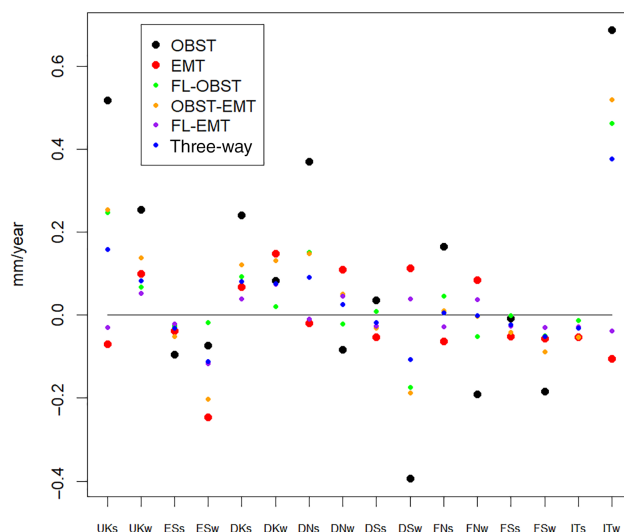

FIGURE 7 Estimates of rainfall trends for the 16 indices based on observations (OBST), the climate model ensemble mean (EMT) and four model combinations. The FL model estimate of the trend is given by the zero line

14 of 17

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

JEWSON ET AL.

In all 16 cases, the three-way trend is smaller than the observed trend. This is because the observed trend is uncertain, and our combination method downweights it significantly as it constructs optimal blended estimates of the size of the trend. In 10 of 16 cases, the three-way model has the same sign as the EMT model because of the relatively large weights on that model.

6 | SENSITIVITY TESTS

We now look at four sensitivity tests to see how much the results from the previous section change, as we change a number of aspects of the methodology. The results are shown in Table 2 and Figure 8 and discussed in Section 6.5.

6.1 | Absolute versus relative trends

In the above analysis, we have taken the trends from the climate model as absolute values in mm/year. However, given that the mean rainfall of the climate model is biased, to some extent, in all locations, one might argue that it makes more sense to take these trends as relative values. In fact, the question of whether to use absolute or relative values is rather subtle, and which makes most sense depends on the source of the bias. If the bias is due to errors in the climate model, it makes most sense to use the relative trend. If, on the other hand, the bias is due to the observations, it makes more sense to use the absolute trend. It may, in fact, be reasonable to use something in-between those two extremes, since the bias undoubtedly

comes from both data sources, to some extent, although we do not follow that approach in this study.

To illustrate the difference between absolute and relative changes: for our Spanish summer mean rainfall example, we find that the mean rainfall in the climate model is on average around 30% too high. One could interpret this to suggest that the trend from the climate model will also be 30% too high.

The impact on the final predictions of the change from an absolute to a relative trend approach is caused by the changing trend slope in the EMT model, a change in the AICC value for the EMT model, and changing weights for all models, since the three weights must adjust so that they still sum to one. The change in the trend slope and the changing weights both affect the final prediction, and analysis of individual cases from our set of 16 shows that they may work in the same or opposite directions.

Relative trend-based estimates are given in Table 1.

6.2 | AIC versus BIC

Our model weighting methodology is based on the use of the AIC. However, there is a commonly used alternative to AIC, known as the Bayesian information criterion (BIC). We recalculate all the results using BIC rather than AIC. In our situation, BIC tends to put more weight on FL and EMT, and less weight on OBST. BIC values for the Spanish summer rainfall example are given in Table 1.

6.3 | Ensemble mean uncertainty

One characteristic of the AICC model combination method is that it ignores the uncertainty on the climate model trend, as if the climate model trend were generated from an infinitely sized ensemble. This is not completely appropriate: the uncertainty on the EMT is certainly much less than the uncertainty on the observed trend (see Figure 3a,b), but it is not zero. Taking account of the uncertainty on the EMT is difficult in the maximum likelihood framework. To explore the impact of this assumption, we rebuild all three models (FL, OBST and EMT) and repeat the three-way combination, using Bayesian principles. The benefit of using Bayesian formulations of the models is that all the estimated parameters and the EMT are now represented by distributions, thus capturing their uncertainty. The disadvantage is that the calculations are much more complex.

A crucial aspect of any Bayesian model is the choice of the prior. For the Bayesian versions of the FL and
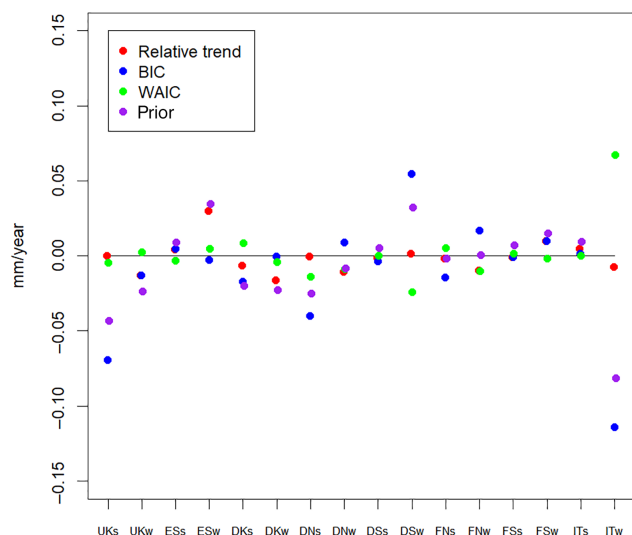

FIGURE 8 Changes in estimates of rainfall trends relative to the basic three-way model for four sensitivity tests

OBST models, we use the standard objective prior formulations for these models. FL and OBST are two of a small number of statistical models that have a unique and more or less uncontroversial objective Bayesian formulation that avoids the need to introduce subjective priors (see, e.g., the discussion in Lee, 1997).

For the Bayesian version of the EMT model, we use the same Bayesian model as for OBST, but now for a given trend. The given trend is sampled from a distribution of trend estimates implied by the spread of the trends from the climate model ensemble, thus capturing the uncertainty on the climate model trend.

For the model comparison between these Bayesian models, AIC theory can no longer be used since it relies on the assumption that models are fitted using maximum likelihood. Instead, we use the widely accepted information criterion theory (WAIC) developed by Watanabe (2010). In this theory, AICC scores are replaced by WAIC scores. WAIC scores are then converted to weights, exactly as in the AIC theory.

In the results from this test, less weight is always placed on the climate model, as would be expected since we are introducing uncertainty around the climate model trend. For smaller ensembles, this effect would become more important.

WAIC values for the Spanish summer rainfall example are given in Table 1.

6.4 | Equal versus unequal prior weights

In the two- and three-way combinations considered above, we implicitly put equal prior weights on the different models. This means that the relative weights on the models are determined solely by their AICC score. However, these implicit prior weights can be seen as a subjective choice. For instance, in the three-way combination, one can argue that because two of the models involve modelling trends, and only one not, we implicitly put twice as much prior weight on the idea that there is a trend than the idea that there is no trend. It could then be argued that this is inappropriate and biases the results towards favouring using a trend.

The question of what prior weights to put on the models can never be resolved completely satisfactorily. It is an example of a general feature of all mathematical modelling, which is that the modelling decisions made by the mathematical modeller inevitably have an impact on the final results. This problem is made worse when the data are insufficient to distinguish completely clearly between the models, as is the case here.

As a sensitivity test, we have put prior weights on the trend models that are half the prior weight put on the FL model, so that the total prior weights are equal on trend and no-trend models.

6.5 | Discussion of sensitivity test results

Of the four sensitivity tests, all had some impact on the results, and the impact varied by index across our 16 indices. The impact of using a relative trend, and switching to a Bayesian formulation in order to include uncertainty on the climate model trend, was relatively small. The impact of using a different prior was larger, and the impact of using BIC, instead of AIC, was largest.

7 | SUMMARY AND CONCLUSIONS

We have studied how to make estimates, or predictions, of current and near-future mean rainfall in Europe, based on estimating the climate change trend in the recent past. This trend can be estimated either from observations or from climate models, or it can be assumed to be so close to zero and so hard to estimate that it is better set to zero.

We have investigated a method that constructs an optimal combination of three separate trend estimates. The three estimates are the flat-line (FL) model, which uses no trend, an observed trend (OBST) model that fits an ordinary least squares (OLS) trend to the data, and the ensemble mean trend (EMT) model that fits an OLS trend to the climate model ensemble mean and uses that trend to model the observed data. The method draws on standard model averaging theory involving the Akaike information criterion (AIC) score. The method takes into account how well the trend can be estimated in the observational data, and to what extent the climate model trend is consistent with the observations. It puts weights on the three models based on how close each model is estimated to be from the unknown truth.

We find that overall the OBST gets the lowest weights, because the observed trends are relatively weak and the estimates of the trends are uncertain. In many cases, the three-way combination reduces the observed trend estimate to less than 20% of its OLS value.

The FL and EMT methods are given higher weights than the OBST method in most cases. That EMT often gets higher weights than OBST demonstrates the value of climate model ensembles for this application. The FL and EMT weights are roughly the same overall, although this varies from case to case. This suggests that overall the EMT method does not fit the data significantly better than the FL method. In particular cases, the EMT method is given slightly higher weights, when the climate

model trend agrees with the observed trend. In cases where FL and EMT are given roughly the same weight, we can see them as different points of view on how to model the trend, based on different assumptions. They cannot be separated by the statistical analysis since they are equally consistent with the data.

Sensitivity tests indicate that the results are somewhat sensitive to the details of the methodology, which indicates that careful thought needs to be given to the various methodological choices.

One of the modelling decisions made in our study was to use the normal distribution to model the residuals around the rainfall trends. We chose to use the normal distribution since this makes the analysis particularly simple, and the residuals are indeed close to normally distributed. However, there might be advantages in using other distributions. For instance, since rainfall residuals are generally positively skewed, and cannot go below zero, there would possibly be advantages to using the log-normal distribution, which can be implemented by taking the log of all the rainfall values before the analysis. The disadvantages of this approach are that it would make the mathematical expressions more complex, and there is no longer a single easily-interpreted value for the trend being estimated.

In this particular application of the model averaging methodology that we have described, most of the indices that we have studied are close to one corner of parameter space where the observational trend gets relatively little weight and the climate model trend and no-trend methods get roughly equal weight. Other data sets for other variables, including other rainfall indices such as extreme rainfall, may well lie in different parts of parameter space. This may lead to different weighting behaviour, depending on the sizes of the trends in the observations and the climate model, the uncertainty on the observed trends and the level of consistency between the climate model and the observations.

The method we have described is a general method for extracting information from observations and climate model simulations of recent climate. The method avoids having to decide whether to model a trend or not, and avoids having to decide whether to believe observed trends or climate model trends. Instead, it uses an objective analysis of the evidence to create a single best combined estimate of the climate change trend. We believe that there is great potential to improve estimates of current climate by including information from climate models in this way.

In the precise form presented here, the combination method only applies to linear trends and normally distributed residuals. However, with modifications, it could equally well be applied to other shapes of trend and other shapes of distribution. Applying similar methods to trends in extremes would also be useful, but would require enhancements to the statistical methodology that would need to go beyond the current literature on model averaging, and hence would require significant research.

CONFLICT OF INTEREST
None.

AUTHOR CONTRIBUTIONS
Stephen Jewson: Conceptualization (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); software (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Francesco Comola:** Data curation (equal); writing – review and editing (equal). **Ben Parkes:** Data curation (lead); writing – review and editing (supporting).

ORCID
Stephen Jewson https://orcid.org/0000-0002-6011-6262

REFERENCES

Birch, C., Parker, D., Marsham, J., Copsey, D. & Garcia-Carreras, L. (2014) A seamless assessment of the role of convection in the water cycle of the west African monsoon. *Journal of Geophysical Research – Atmospheres*, 119, 2890–2912.

Burnham, K. & Anderson, D. (2002) *Model selection and multimodel inference.* New York: Springer-Verlag.

Claeskens, G. & Hjort, N. (2008) *Model selection and model averaging.* Cambridge: CUP.

Cornes, R., van der Schrier, G., van den Besselaar, E. & Jones, P. (2018) An ensemble version of the E-OBS temperature and precipitation datasets. *Journal of Geophysical Research – Atmospheres*, 123(17), 9391–9409.

Fletcher, D. (2019) *Model averaging.* Berlin: Springer.

Friedman, D. (1972) Insurance and the natural hazards. *ASTIN*, 7, 4–58.

Garcia-Carreras, L., Challinor, A.J., Parkes, B.J., Birch, C.E., Nicklin, K.J. & Parker, D.J. (2015) The impact of parametrized convection on the simulation of crop processes. *Journal of Applied Meteorology and Climatology*, 54, 1283–1296.

Grossi, P. & Kunreuther, H. (2005) *Catastrophe modelling: a new approach to managing risk.* New York: Springer.

Hall, T. & Jewson, S. (2007) Statistical modelling of North Atlantic tropical cyclone tracks. *Tellus*, 59A, 486–498.

Herrera, S., Kotlarski, S., Soares, P.M.M., Cardoso, R.M., Jaczewski, A., Gutiérrez, J.M. et al. (2019) Uncertainty in gridded precipitation products: influence of station density, interpolation method and grid resolution. *International Journal of Climatology*, 39, 3717–3729.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M. et al. (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14, 563–578.

Jewson, S., Brix, A. & Ziehmann, C. (2005) *Weather derivative valuation*. Cambridge: CUP.

Jewson, S., Dallafior, T. & Comola, F. (2021) Dealing with trend uncertainty in empirical estimates of European rainfall climate for insurance risk management. *Meteorological Applications* (in press). 28, e2008.

Kaczmarska, J., Jewson, S. & Bellone, E. (2018) Quantifying the sources of simulation uncertainty in natural catastrophe models. *SERRA*, 32(3), 591–605.

Knutson, T. & Tuleya, R. (2004) Impact of CO_2-induced warming on simulated hurricane intensity and precipitation: sensitivity to the choice of climate model and convective parametrisation. *Journal of Climate*, 17, 3477–3495.

Konduru, R. & Takahashi, H. (2020) Effects of convection representation and model resolution on diurnal precipitation cycle over the Indian monsoon region: toward a convection-permitting regional climate simulation. *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032150.

Lee, P. (1997) *Bayesian statistics*, 2nd edition. London: Arnold.

Loridan, T., Scherer, E., Dixon, M., Bellone, E., Khare, S. et al. (2013) Cyclone wind field asymmetries during extratropical transition in the western North Pacific. *Journal of Applied Meteorology and Climatology*, 53, 421–428.

Lowe, J., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D. et al. (2018) UKCP science overview report [Online]. Available at: https://www.metoffice.gov.uk/research/collaboration/ukcp/guidance-science-reports. Accessed September 9, 2021.

Maraun, D. (2012) Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. *Geophysical Research Letters*, 39, L06706.

Maraun, D. (2013) When will trends in European mean and heavy daily precipitation emerge? *Environmental Research Letters*, 8(1), 1–7.

Maraun, D. & Widmann, M. (2018) *Statistical downscaling and bias correction for climate research*. Cambridge: CUP.

Michel, G. (2018) , *Risk modelling for hazards and disasters*. Amsterdam: Elsevier.

Mitchell-Wallace, K., Jones, M., Hillier, J. & Foote, M. (2017) *Natural catastrophe risk management and modelling*. Chichester: Wiley Blackwell.

Risbey, J., Pook, M. & McIntosh, P. (2013) Spatial trends in synoptic rainfall in southern Australia. *GRL*, 40(14), 3781–3785.

Sassi, M., Nicotina, L., Pall, P., Stone, D., Hilberts, A., Wehner, M. et al. (2019) Impact of climate change on European winter and summer flood losses. *Advances in Water Resources*, 129, 165–177.

von Storch, H. & Zwiers, F. (1999) *Statistical analysis in climate research*. Cambridge: CUP.

Vickery, P., Skerlj, P., Lin, J., Twisdale, L., Young, M., Lavelle, F. et al. (2006) Hazus-MH hurricane model methodology. II: damage and loss estimation. *Natural Hazards Review*, 7(2), 94–103.

Wasserman, L. (2004) *All of statistics*. New York: Springer.

Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(1), 3571–3594.

Wilks, D. (2011) *Statistical methods in the atmospheric sciences*, 3rd edition. Oxford: AP.

Willetts, P., Marsham, J., Birch, C., Parker, D., Webster, S., Petch, J. et al. (2017) Moist convection and its upscale effects in simulations of the Indian monsoon with explicit and parametrized convection. *Quarterly Journal of the Royal Meteorological Society*, 143, 1073–1085.