Model skill measures in probabilistic regional climate projections for Ireland

3 authors:

Aideen M Foley
Birkbeck, University of London
20 PUBLICATIONS **252** CITATIONS

SEE PROFILE

Rowan Fealy
National University of Ireland, Maynooth
55 PUBLICATIONS **751** CITATIONS

SEE PROFILE

John C. Sweeney
National University of Ireland, Maynooth
84 PUBLICATIONS **1,087** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Revision of potato late blight forecast in Ireland. View project

Project   EURO-CORDEX View project

Model skill measures in probabilistic regional climate projections for Ireland

Aideen Foley[1,2,*], Rowan Fealy[1], John Sweeney[1]

[1]Department of Geography, NUI Maynooth, Co. Kildare, Ireland

[2]Present address: Department of Land Economy, University of Cambridge, Cambridge CB3 9EP, UK

ABSTRACT: In the present study, a range of regional climate models have been used to test approaches to Bayesian model averaging (BMA), particularly the quantification of model weights/Bayesian priors. The results of skill assessments were used to inform probabilistic future projections of Irish climate using a BMA approach in order to evaluate how different approaches to skill assessment, based on representation of climate means, or of a large-scale driver (the NAO), or a combination thereof, may influence the final climate projection. Results indicate that means-based skill assessments may not always provide a useful indication of model skill and that further analyses are required to assess a model′s ability to simulate the dynamics of the climate system. While this research illustrates that the use of metrics derived from the model predicted NAO impacts on the regional projection, it also supports the inclusion of other large-scale model diagnostics. When used to weight model projections to produce ensemble climate projections, the choice of skill metric may have an impact on the shape of the probability distribution and the most probable outcome of future climate predictions. The present study demonstrates that when working with probabilistic outputs of ensemble climate modelling experiments, awareness of the approaches used to evaluate models and the techniques used to combine them to formulate ensemble projections are integral in enabling robust responses to the potential changes in climate represented by models.

KEY WORDS: Regional climate modelling · Model assessment · Ireland · NAO · Bayesian model averaging

1. INTRODUCTION

When developing future climate projections, model skill is a key consideration. Models in which skill is derived from error cancellation cannot be relied upon under different forcings. As such, a model that skilfully captures large-scale drivers of climate in a region is likely to be a more robust tool. The aim of this research was to assess how projections of future climate are altered when different levels of knowledge about model skill are incorporated into the projection generation process, using the Irish domain as a case study. This is a region that has received little attention in regional climate modelling (RCM) literature, making the projections developed in the pres-ent study of key importance for future planning in this area, although the conclusions drawn are also relevant to all such studies.

In addition to widely used metrics, such as correlation and bias, projection development also takes account of the ability of RCMs to simulate the climatic behaviour associated with the North Atlantic Oscillation (NAO). Due to the importance of the NAO to climate variability in the North Atlantic region, the impacts associated with its phases and the potential for climate change to alter its behaviour, it is desirable that models are able to represent this mode of variability. For this reason, correlation and bias are presented briefly in this paper, with the major focus on representation of the NAO.

To determine the difference in projections when varying levels of knowledge are used to inform model choice, a Bayesian model averaging (BMA) approach was used with the uncertainty surrounding each model reflected in the weights associated with each scheme. Bayesian approaches have the advantage of accounting for uncertainty in model selection, thus reducing the potential for over-confident projections (Hoeting et al. 1999, Harris et al. 2010). In this research, various weighting schemes were used, informed to varying degrees about model performance. The outcomes obtained using each method are compared and discussed.

2. APPROACHES TO GENERATING ENSEMBLE CLIMATE MODEL PROJECTIONS

Simple ensemble methods have a history of use within short-term weather forecasting, and a widely used approach is to treat the ensemble mean as a best estimate of future conditions (Whitaker & Loughe 1998). Examples of the mean ensemble method applied to climate model data include Gates et al. (1999) and Rinke et al. (2006). Both of these studies found that ensemble means outperform individual models for certain climate parameters. Probabilistic ensemble methods are an alternative approach, and can provide a better understanding of potential future conditions compared with deterministic or simple ensemble approaches (e.g. New et al. 2007, Lopez et al. 2006, Bouwer et al. 2010).

However, Tracton & Kalnay (1993) noted that the increase in skill associated with ensemble methods may in part be due to the cancellation of errors in the individual forecasts when the ensemble members are averaged. If one model is particularly lacking in simulative skill, the ensemble mean forecast will be affected by this; thus, differences in skill should be considered when formulating ensemble projections (Grimit & Mass 2002). Much information can be generated about model skill through comparison with present-day observational data. Incorporating this information in ensemble projection provides an opportunity to account for uncertainty and increase confidence in the ensemble, provided, of course, that the models are all skilful to some degree.

Examples of weighted multi-model ensemble studies include Yun et al. (2005), Sanchez et al. (2009) and Coppola et al. (2010). Weighting systems often rely on skill scores calculated by comparing modelled and observed climate parameters. However, as Brown (2004) noted, simply comparing climate model outputs can result in model skill being under- or overestimated. Skill in representing the mean field or a single key climate parameter does not guarantee that the processes and drivers that give rise to mean temperature or precipitation, for example, are adequately represented in the model. Lucarini et al. (2007) noted that the focus on mean fields has greatly influenced the development of general circulation models (GCMs) and suggested that, as the climate system is essentially a non-linear system, it would be more appropriate for model validation to include analysis of the representation of dynamical processes.

Another approach is to weight models according to their relative agreement with each other (Giorgi & Mearns 2003, Tebaldi et al. 2004). However, as noted by Abramowitz (2010), model independence is rarely quantified, and shared characteristics may lead to a high degree of similarity between models. Therefore, in some cases, weighting by relative agreement may result in overconfidence in the outcomes. The application of ensemble methods is often understood to generate an increase in reliability (Tebaldi & Knutti 2007), but this reliability is optimised when ensemble members are independent and more limited when they are not.

There are some examples in the literature of weighting approaches based on the models′ abilities to simulate the dynamics of the climate system rather than the mean fields. For example, Schmittner et al. (2005) used model skill in representing key hydrographic properties and circulation estimates to weight members in an atmosphere–ocean coupled GCM (AOGCM) ensemble. Yet the mean-based skill scores approach remains the more widely used technique. However, there may be an opportunity to reduce the uncertainty associated with future projections and improve their reliability by incorporating information about how key climate drivers are represented in the models. Of course, as with any model evaluation, there is an element of subjectivity associated with the choice of metrics, and in any attempt to constrain future projections subjective decisions must be made regarding how information should be combined to form weights (Christensen et al. 2010).

3. DATA AND METHODOLOGY

3.1. Data and model selection

RCM output for 1961–1990, all at 0.5 by 0.5° resolution, were obtained from the PRUDENCE data

archive (Table 1). Gridded observational data made available via the British Irish Council were used to evaluate temperature and precipitation (British Irish Council 2003). Mean sea level pressure (MSLP) fields from the ERA-40 reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts were also used. The ERA-40 data were interpolated from their native grid to the finer grid used by the RCMs for comparative purposes. RCM output for 2071–2100 were also obtained from the PRUDENCE data archive. ARPEGE, which was forced using observed sea-surface temperatures (SSTs) in the control period, was forced using the HadCM3 AOGCM for the period 2071–2100. HadRM3P, HIRHAM and ARPEGE, which were each run 3 times to create sub-ensembles for the control and A2 scenarios, were only run once for the B2 scenario. Therefore, there were more simulations available for the A2 scenario than for the B2 scenario.

However, definite similarities within GCM driver groups occurred in the PRUDENCE ensemble, illustrated in Fig. 1. Errors in the driving GCMs (e.g. Jacob et al. 2007) may propagate through to the RCMs shown in Fig. 1, potentially leading to similar effects on temperature and precipitation. Therefore, a subset of these models, representative of different combinations, was selected to develop probabilistic climate projections for Ireland, as including models driven by the same GCM could result in over-confidence. The RCMs chosen were HadRM3P-a, RCAO-H, HIRHAM-E5, ARPEGE-a and RCAO-E4.

3.2. Investigation of the relationship between model NAO error and simulated regional climate

To assess whether errors in representation of the NAO affect the simulation of regional climate, an automated Lamb classification (Jenkinson & Collison 1977) was used to classify monthly MSLP data. Applications of this technique have been published by Goodess & Palutikof (1998) and Linderson (2001).

Positive NAO phases are associated with increased westerly winds in winter and, correspondingly, with higher levels of precipitation. To determine whether these characteristics were simulated by the models, the frequencies with which the various wind directions occurred and the precipitation amounts associated with each classification were extracted from the data using the following technique.

Equations were calculated using the points indicated in Fig. 2, to determine the predominant wind direction in each winter month for both the modelled and observational data:

$$W = \frac{1}{4}(7+8+9+10) - \frac{1}{4}(1+2+3+4)$$

$$S = \frac{1}{\cos(53.5°)}\left(\left\{\frac{1}{4}[4+10+(2\times6)]\right\} - \left\{\frac{1}{4}[1+7+(2\times5)]\right\}\right) \quad (1)$$

$$D = \tan^{-1}(W/S)$$

where W is westerly flow, S is southerly flow, D is wind direction and 53.5° is the bisecting latitude of the analysis grid (Fig. 2). If W was positive and S was negative, 360° was added to D; in all other cases, 180°

Table 1. Summary of models available from the PRUDENCE archive. Models with both A2 and B2 scenario data are denoted by asterisk; models chosen for scenario development are in italics. GCM: general circulation model; RCM: regional climate model; SST: sea-surface temperature

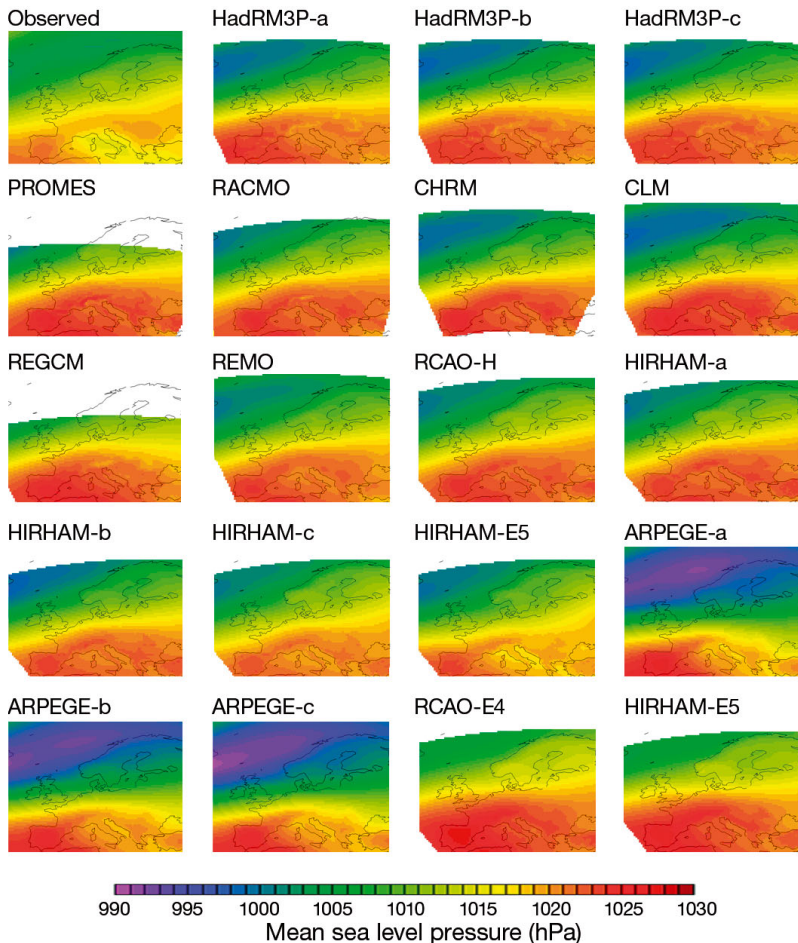| GCM	RCM	Source
HadCM3/HadAM3P; Pope et al. (2000)	*HadRM3P-a**, HadRM3P-b, HadRM3P-c	Moberg & Jones (2004)
HadCM3/HadAM3H; Pope et al. (2000)	PROMES*	Castro et al. (1993)
	RACMO	Räisänen et al. (2004)
	CHRM	Vidale et al. (2003)
	CLM	Doms & Schlatter (2002)
	REGCM*	Giorgi et al. (1993)
	REMO	Jacob & Podzun (1997)
	*RCAO-H**	Döscher et al. (2002)
	HIRHAM-a*, HIRHAM-b, HIRHAM-c	Christensen et al. (1996)
ECHAM4-OPYC/ECHAM5; Roeckner et al. (1996)	*HIRHAM-E5*	
Observed SSTs	*ARPEGE-a**, ARPEGE-b, ARPEGE-c	Déqué et al. (1998)
ECHAM4-OPYC; Roeckner et al. (1996)	*RCAO-E4**	
	HIRHAM-E4*	

Observed HadRM3P-a HadRM3P-b HadRM3P-c
PROMES RACMO CHRM CLM
REGCM REMO RCAO-H HIRHAM-a
HIRHAM-b HIRHAM-c HIRHAM-E5 ARPEGE-a
ARPEGE-b ARPEGE-c RCAO-E4 HIRHAM-E5

990   995   1000   1005   1010   1015   1020   1025   1030
Mean sea level pressure (hPa)

Fig. 1. Modelled mean sea level pressure (hPa) for winter, 1961–1990

the different measures of present-day skill. For values within the combined range of all the models, the likelihood for each model was calculated. The likelihood associated with each model was the probability density associated with the climate variable value for a normal distribution specified using the mean and standard deviation of that model.

Normality of 30 yr seasonal data-sets of both temperature and precipitation in the control period was tested using the Shapiro-Wilks test. For both parameters, the data were found, for the most part, to be normally distributed. Additionally, according to the central limit theorem, it is reasonable to use the normal distribution to generate likelihood functions for future seasonal temperature and precipitation. These likelihoods are then multiplied by the respective weights to form the posterior distribution. In this case, the posterior was a weighted ensemble probability density function (PDF), which took account of intermodel uncertainty and information about model skill such that:

$$p(x \mid x_1, ..., x_N, x^T) = \sum_{n=1}^{N} w_n g_n(x \mid \bar{x}_n, \sigma^2) \qquad (2)$$

where $p(x \mid x_1, ..., x_N, x^T)$ is the ensemble PDF for the climate parameter x, given projections from n models $x_1, ..., x_N$ and present-day data x^T, w_n is the weight for each model and $g_n(x \mid \bar{x}_n, \sigma^2)$ is a theoretical normal PDF for each model defined by mean $\bar{x}_n$ and variance σ^2 from the future projections of each model n.

was added. Equations were adapted from Jones et al. (1993). Point references are italic.

Although wind direction can vary significantly on a daily basis, comparisons of monthly data from RCAO driven by ECHAM4 and daily data from the ECHAM4 GCM show that applying the method to monthly data adequately captures the overall shape of the frequency distribution for wind direction.

3.3. Bayesian model averaging

BMA was used to generate climate projections over land for Ireland, utilising a range of skill-scoring methods. Bayesian statistics differ from frequentist statistics in that subjective information regarding the 'level of knowledge' about projections can be incorporated into the ensemble projection through the use of an informative prior. In this case, the priors were weights, determined based on

3.4. Skill metrics used in Bayesian priors

The priors used in the application of BMA were weights derived for each model, based on a selection of skill metrics. Several weighting schemes were developed (Fig. 3). In the simplest scheme (BMA-EQ), models were assumed to be equally skilful and assigned equal weights.

A second scheme based on a selection of traditional skill-scores (BMA-SS) was also developed. For these
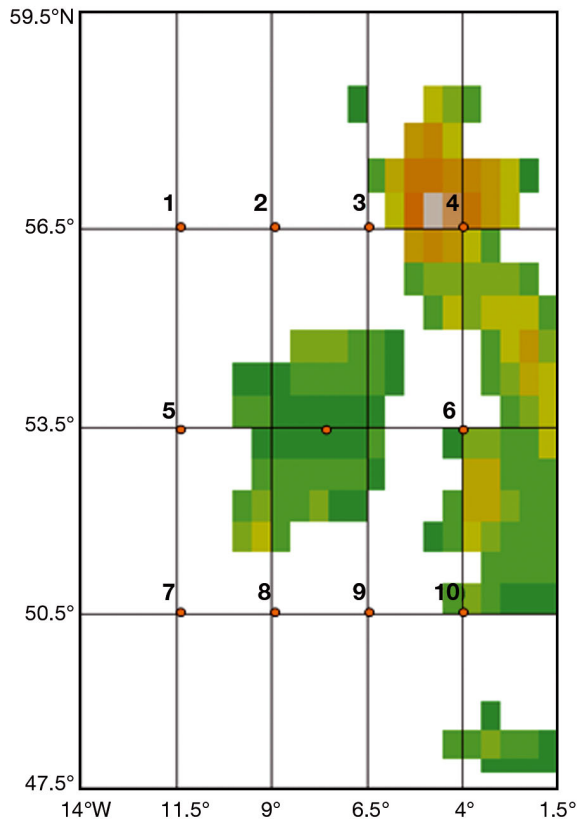
Fig. 2. Points used to determine predominant wind direction

weights, model skill was quantified using the following metrics. Mean annual climatologies were calculated for each model based on their 20th century hindcasts and were compared to a mean annual climatology based on observational data for the same period. Agreement was quantified by calculating the correlation coefficient. Additionally, mean seasonal spatial patterns for each season were extracted from both the model hindcasts and observational data, and agreement was quantified using the correlation coefficient. The temporal correlation coefficient based on mean annual climatologies was combined with appropriate spatial correlation coefficients based on mean seasonal spatial patterns, depending on the season for which probabilistic projections were being developed, to form the weight or prior for this weighting scheme. The metrics were averaged and then squared so that low skill was more heavily weighted against and normalized so that the metrics resulted in a sum of 1.

A third scheme was based on skill in simulating the NAO, a key large-scale driver of Irish climate. Studies have demonstrated links between NAO and Irish climate (Kiely 1999, Wilby et al. 1997), but the concept could also be applied to other large-scale drivers.

The observed NAO index could not be used to identify NAO-positive and -negative (NAO+ and NAO−, respectively) years in the models, as differences in the initial and boundary conditions of the models led the individual simulations to evolve differently from observations. Therefore, a model NAO index was calculated for each RCM. As the domain of the RCMs did not include Reykjavik (64° N, 22° W), the closest available point (64° N, 14° W) was used to represent this station. Lisbon (38° N, 9° W) was chosen as the southern point. At each pressure centre, winter MSLP anomalies were calculated and normalized by dividing by the standard deviation. The model NAO index was the difference between the normalized Reykjavik and Lisbon MSLP anomalies. As this method uses different pressure centre points to those commonly used to calculate the NAO index, validity of the method was tested by applying it to ERA-40 MSLP data and comparing results with the NAO index calculated by the Climate Analysis Section at the National Centre for Atmospheric Research. The difference in location was found to make no difference to the calculation.

To eliminate noise, years with an NAO index of between −1 and +1 were omitted. Discretising the NAO in this way, as strong positive and strong negative phases, and analysing the associated climate patterns allows for fast identification of models that possess skill in simulating the climatic behaviour associated with the NAO. This provides for a more
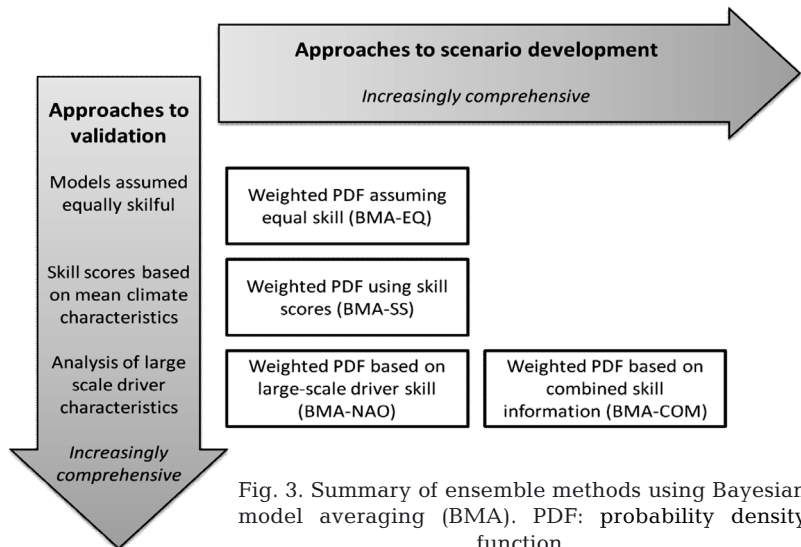

Fig. 3. Summary of ensemble methods using Bayesian model averaging (BMA). PDF: probability density function

stringent test of the climate models, which are known to reproduce mean climate better than the extremes. It also facilitates a more robust comparison between models, as the spatial climate patterns associated with the strongly positive or strongly negative phases of the NAO should be more defined. For MSLP, temperature and precipitation, the mean spatial patterns associated with NAO+ years and NAO− years were mapped, and the level of agreement between observed and modelled patterns was ranked, using a simple, qualitative scheme. Scores of 0.99, 0.66 and 0.33 were assigned based on performance in simulating the NAO, ranging from 0.99 for the most skilful models to 0.33 for the least skilful. This range was chosen to match the 0 to 1 range of the correlation coefficients as it allows the different skill scores to be easily combined. These scores were again squared and normalized so that they summed to 1.

The final scheme (BMA-COM) took both skill scores and NAO performance into account. Normalized NAO skill scores were added to the normalized spatiotemporal skill scores to form a combined objective skill estimate. The objective skill estimates were then normalized to form the BMA weights for this combined approach.

To summarize, weights w for each approach were calculated as follows:

$$\text{BMA-EQ:} \quad w_{EQ} = \frac{1}{n} \qquad (3)$$

where n is the number of models and

$$\text{BMA-SS:} \quad w_{SS} = (\bar{r})^2 \qquad (4)$$

where $r = r_S + r_T$, with r_S being the correlation between modelled and observed mean seasonal spatial patterns and r_T being the correlation between modelled and observed mean annual climatologies.

$$\text{BMA-NAO:} \quad w_{NAO} = (s)^2 \qquad (5)$$

where s is the NAO skill score and

$$\text{BMA-COM:} \quad W_{COM} = (w_{SS} + w_{NAO}). \qquad (6)$$

In all cases, the normalized weight is obtained by:

$$w_i' = \frac{w_i}{\sum_{i=1}^{n} w_i} \qquad (7)$$

where n is the number of models.

The issue of systematic bias was addressed by applying a correction factor to those models that required it, while distinguishing between systematic and random bias. An r-value of 0.7 or higher was regarded as signifying a strong association between the observed and modelled patterns, while values of <0.7 signified weak to moderate association. Models with a bias of >10% of observed precipitation and temperature (0.37 mm d^{-1} or 0.47°C in winter, 0.25 mm d^{-1} or 1.39°C in summer) and which displayed a Pearson r of >0.7 were assumed to be systematically biased and were therefore corrected by subtracting the measured bias based on the present-day simulation from the future value. Spatial r-scores of <0.7 were considered indicative of potentially random bias, and such models were left unchanged, as this bias was less likely to retain the same spatial distribution pattern under different forcing conditions. This approach was assumed to be less subjective than assuming that all model biases are systematic, but it is important to recognize the treatment of data before developing probabilistic projections as a potential source of uncertainty.

4. RELATIONSHIP BETWEEN SIMULATED LARGE-SCALE VARIABILITY AND REGIONAL CLIMATE

Fig. 1 illustrates that many models within the PRUDENCE ensemble have stronger than observed MSLP gradients in winter, with similar patterns emerging from RCMs driven by the same GCMs. A key question is whether this has an effect on the simulated regional Irish climate; this was investigated by exploring the relationship between westerlies and precipitation in the selected models.

Fig. 4 displays observed and modelled wind direction frequencies and associated precipitation amounts for the case study models. Notably, none of the models skilfully simulated the observed wind direction distribution. While HadRM3P-a and RCAO-H displayed a similar wind direction distribution, perhaps due to sharing GCM drivers from the same model centre, the associated precipitation amounts were quite different. Most of the rainfall in these models was associated with south-westerly winds, and this wind direction occurred with similar frequency in both models.

However, in RCAO-H more rain was associated with these winds, making it a wetter model overall. Although HadRM3P-a slightly overestimated south-westerly rain compared to observations, it underestimated rain associated with all other wind directions, resulting in a modelled climate which is drier than observed. In both HIRHAM-E5 driven by ECHAM5 and

RCAO-E4 driven by ECHAM4-OPYC, the difference in frequency between south-westerlies and westerlies was less pronounced. However, more precipitation was associated with these wind directions in RCAO-E4, making it a wetter model. In HIRHAM-E5, although the fractions of precipitation that could be attributed to the different wind directions were not the same as those observed, when totalled they amounted to a similar level of rain, making HIRHAM-E5 apparently a more skilful model overall.

ARPEGE considerably overestimated the frequency of westerlies, and had a correspondingly large amount of associated precipitation. However, while the other models also showed some contributions from the southeast, south and northwest directions, these winds were greatly underestimated in ARPEGE. As a result, these wind directions contributed less to the ARPEGE precipitation total. Therefore, ARPEGE-a appeared to simulate average winter precipitation skilfully, but produced a correct overall precipitation amount in this domain for incorrect reasons.

It is interesting to note that the models driven by HadAM3P, HadAM3H and ECHAM5 have similar spatial patterns of MSLP bias. All 3 of these models are atmosphere-only GCMs, used as part of a double-nested technique to drive the RCMs. This may indicate that, for a maritime country like Ireland, a fully-coupled AOGCM is a better choice of driver. These results demonstrate that errors in representing large-scale drivers can have a significant impact on the simulation of regional climate, providing the motivation for a Bayesian framework which accounts for model skill in simulating the NAO.
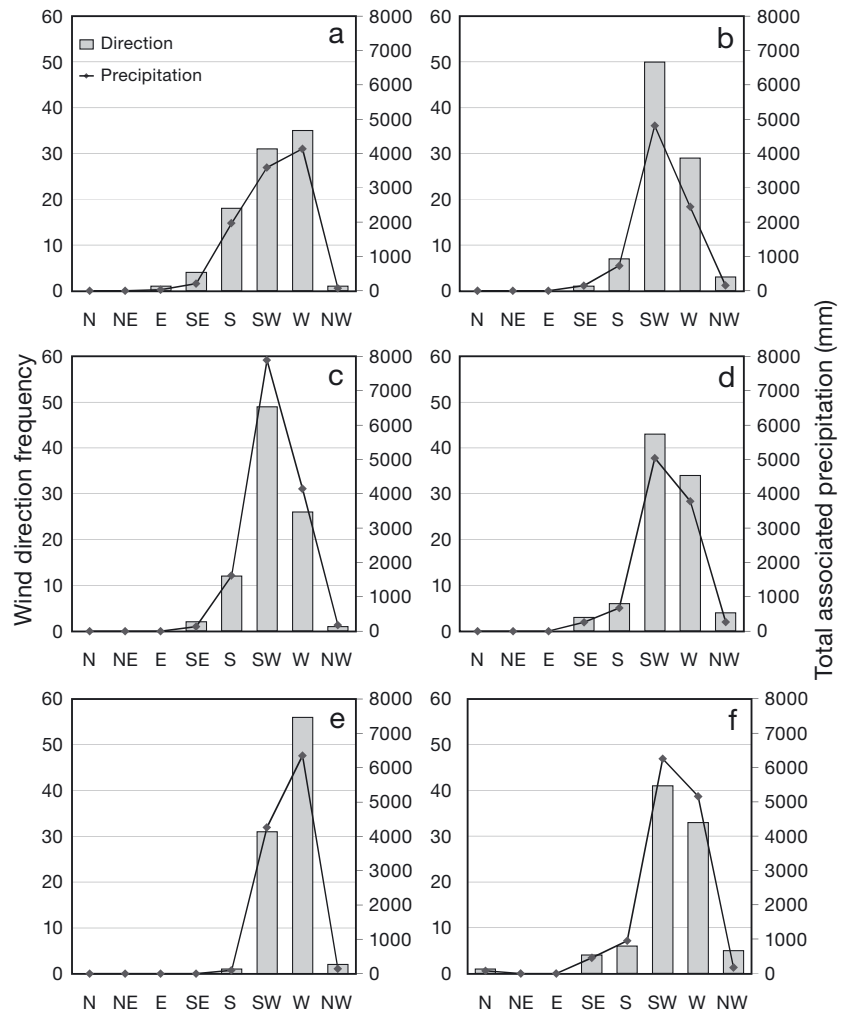

Fig. 4. Observed and modelled wind frequency distributions and precipitation totals from monthly data for (a) observed data, (b) HadRM3P-a, (c) RCAO-H, (d) HIRHAM-E5, (e) ARPEGE-a and (f) RCAO-E4

5. MODEL SKILL IN SIMULATING NAO EFFECTS

MSLP, temperature and precipitation patterns associated with NAO phases are given in Figs. 5, 6 & 7. Observed spatial patterns are provided as an indicator of expected behaviour. In positive NAO years, a distinct pressure gradient was observed, temperatures were warmer and there was more precipitation. The increased precipitation was especially notable in areas which were more exposed to the Atlantic, such as the west coast of Ireland. As the NAO is a significant driver of variability in winter climate in this region, skilful representation of its regional climatic effects would be a very desirable ability in a climate model.

5.1. MSLP

HadRM3P modelled the enhanced pressure gradient associated with a positive NAO phase. However, MSLP in negative NAO years was not as uniform across the domain as in observations. European MSLP, with no NAO division of data, showed a marked negative bias to the north of the domain and
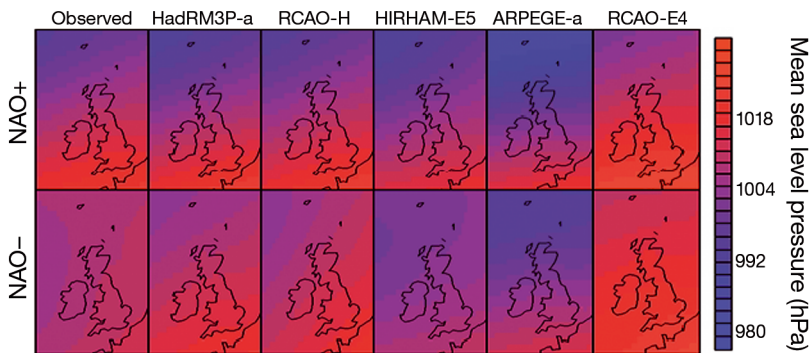
Fig. 5. Observed and modelled mean sea level pressure for the positive and negative phases of the North Atlantic Oscillation (NAO)
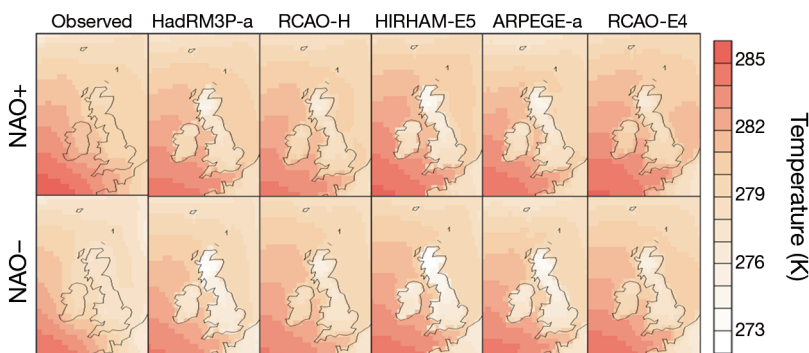

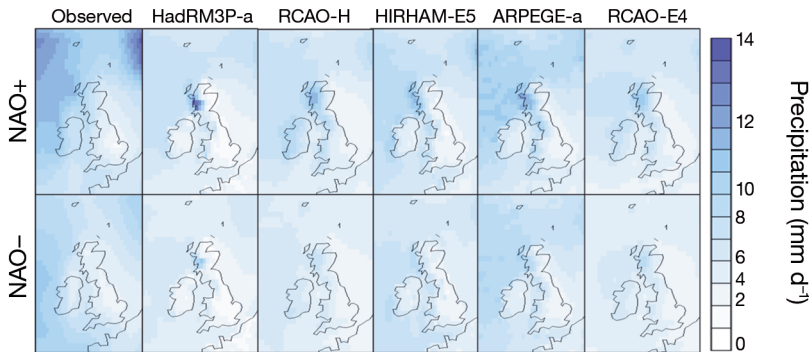Fig. 6. Observed and modelled temperature for the positive and negative phases of the NAO

Fig. 7. Observed and modelled precipitation for the positive and negative phases of the NAO

could explain the gradient associated with average winter MSLP. Although there was a notable bias in HIRHAM-E5's representation of seasonal average MSLP, it captured the difference in pressure patterns between NAO+ and NAO− years quite well. The model appears to capture the pressure differences that underlie the NAO, but it is important to also consider whether these pressure differences have the effect on temperature and precipitation that is expected. ARPEGE-a appears to model a pronounced pressure gradient regardless of NAO phase. Although pressure across the north of the UK and Irish domain was lower in NAO+ years, it was only slightly higher in NAO− years. RCAO-E4 displayed a systematic positive MSLP bias across the European domain, and it is evident from the NAO-related MSLP patterns that pressure in this model was indeed much higher than in the others. However, the model did capture the difference in pressure patterns for positive and negative NAO years, with a north–south gradient occurring in positive years and more uniform conditions occurring in negative years.

5.2. Temperature

Fig. 6 illustrates observed and modelled temperature for the NAO+ and NAO− phases. With an enhanced pressure gradient, one might expect the HadRM3P-a simulation to also model enhanced NAO effects on regional climate. While temperature over the ocean in NAO+ years was warmer than in NAO− years, with warmer temperatures occurring further north, the difference on land was less pronounced. Temperature over both the ocean and land in the RCAO-H simulation exhibited a similar pattern in both NAO+ and NAO− years, with only slight differences in parts of the domain.

In the HIRHAM-E5 simulation, temperature appeared to be warmer in NAO+ years than in NAO− years, and this pattern was especially apparent over

a marked positive bias to the south in HadRM3P-a. Analysis of the UK and Irish domain shows that this error arose through the combination of a slight pressure gradient in NAO− phases and an enhanced gradient in NAO+ phases. RCAO-H also simulated a slight pressure gradient across the UK and Ireland in NAO− years. In NAO+ phases, the pressure gradient was quite pronounced, and, in combination, this

ocean grid cells. In ARPEGE-a, there was minimal difference between the temperature patterns associated with NAO+ and NAO– years in this model, which could be expected due to the similarities between MSLP patterns in NAO+ and NAO– years.

In the RCAO-E4 simulation, there was a notable difference between temperature patterns associated with NAO+ and NAO– years, with NAO+ years being simulated as warmer. Additionally, this warming was more apparent over land grid cells than in some other models.

5.3. Precipitation

Fig. 7 illustrates observed and modelled precipitation for the NAO+ and NAO– phases. HadRM3P-models increased precipitation in the NAO+ phase over western Scotland; however, in the observed NAO patterns the increased precipitation was a domain-wide characteristic and not restricted to this specific area. For Ireland, there was minimal difference between positive and negative NAO years. RCAO-H captured the increased precipitation of the NAO+ phase well, particularly over western Ireland and Scotland (Fig. 6). Since the pressure gradient was enhanced in RCAO-H and the model captured the effects of NAO activity on precipitation with skill, the wetter than observed winter conditions simulated by this model may be attributable to the over-enhanced representation of the NAO. HIRHAM-E5 models increased precipitation in the NAO+ phase, particularly over western Ireland and Scotland (Fig. 6). These results suggest that HIRHAM-E5 captured the NAO and its associated effects on regional climate reasonably well. However, if the NAO was captured accurately, the large biases in MSLP found in this model should have led to an amplification of NAO behaviour, yet this did not occurred here.

In ARPEGE-a, NAO+ years tended towards wetter conditions, particularly over western Scotland. This model produced a significant mean winter pressure gradient bias, but little preservation of the temperature patterns associated with NAO phases at the regional scale.

In addition to capturing the effects of NAO on MSLP and temperature, RCAO-E4 also simulated an effect on precipitation, with NAO+ years tending towards wetter conditions. Although this model simulated erroneous values for mean temperature and precipitation, it captured the dynamics of the NAO quite well. While the systematic pressure bias should not interfere with the gradient of pressure across the

domain, further analysis may indicate whether this error results in an amplified NAO effect, which could in turn explain the systematic errors in temperature and precipitation.

6. PROJECTIONS OF FUTURE IRISH CLIMATE USING BMA APPROACHES

6.1. Formulation of skill scores

Skill levels associated with the models' abilities to represent characteristics of the NAO and characteristics of mean climate are summarized in Table 2. In some cases, such as RCAO for winter temperature, the highest r-values were also accompanied by the largest biases. Other models simulated both large biases and low r-values for certain parameters, for example, summer precipitation in HadRM3P-a. No single model emerged as being skilful in every respect. This information about model skill was subsequently used to weight model projections of Irish climate over land.

HIRHAM-E5 and RCAO-E4 were found to be the most skilful, capturing the MSLP, temperature and precipitation patterns associated with NAO behaviour. RCAO-H and ARPEGE-a captured the precipitation effects of the NAO, but only slight temperature differences and incorrect MSLP patterns. Finally, HadRM3P-a did not capture the precipitation or MSLP patterns associated with NAO behaviour for Ireland adequately and only simulated a slight temperature difference between NAO+ and NAO– years. Scores of 0.99, 0.66 and 0.33 were assigned based on performance in simulating the NAO, ranging from 0.99 for the most skilful models to 0.33 for the least skilful.

6.2. Winter temperature projections (2071–2100)

Projections for winter temperature under the A2 and B2 scenarios are given in Fig. 8. Using BMA approaches, the most likely temperature value for winter temperature under the A2 emissions scenario for 2071–2100 fell between 6.8 and 7°C. Although all approaches yielded a similar 'most likely' value, the data underlying the averages changed significantly when different weightings were used, with different models emerging as the most influential in each weighting scheme. For example, HIRHAM-E5 was more influential when BMA-NAO weighting was used compared with BMA-SS weighting, as

Table 2. Skill estimates based on bias and correlation metrics, and analysis of the model-simulated North Atlantic Oscillation (NAO) in the control period. MSLP: mean sea level pressure; DJF: winter; JJA: summer

	North Atlantic Oscillation			Temperature				Precipitation			
	NAO+/- spatial patterns			Seasonal mean % bias		Seasonal spatial r		Seasonal mean % bias		Seasonal spatial r	
	MSLP	Temperature	Precipitation	DJF	JJA	DJF	JJA	DJF	JJA	DJF	JJA
HadRM3P-a	Low skill	Moderate skill	Low skill	22	2	0.57	0.73	17	34	0.77	0.42
RCAO-H	Low skill	Moderate skill	High skill	40	1	0.88	0.64	40	23	0.78	0.64
HIRHAM-E5	High skill	High skill	High skill	6	4	0.48	0.65	0	7	0.84	0.62
ARPEGE-a	Low skill	Moderate skill	High skill	11	2	0.60	0.77	9	4	0.72	0.22
RCAO-E4	High skill	High skill	High skill	53	1	0.91	0.54	29	25	0.73	0.60

Low skill ▭▭▭▭▭▭▭▭▭▭ High skill

HIRHAM-E5 had a low spatiotemporal skill score, but a high NAO skill estimate. Conversely, HadRM3P-a had a higher weight under BMA-SS weighting than it did under BMA-NAO weighting. Due to these differences, the contribution of each model to the ensemble PDF varied under each weighting scheme. Though similar results can be obtained for a mean projection even when the underlying data vary, if there is to be confidence in the mean projection, the underlying data must be assessed and combined according to the relative merits and deficiencies of the models.

Under the B2 emissions scenario, as no HIRHAM-E5 B2 data were available, only 4 simulations were used, changing the weights and contributions of each
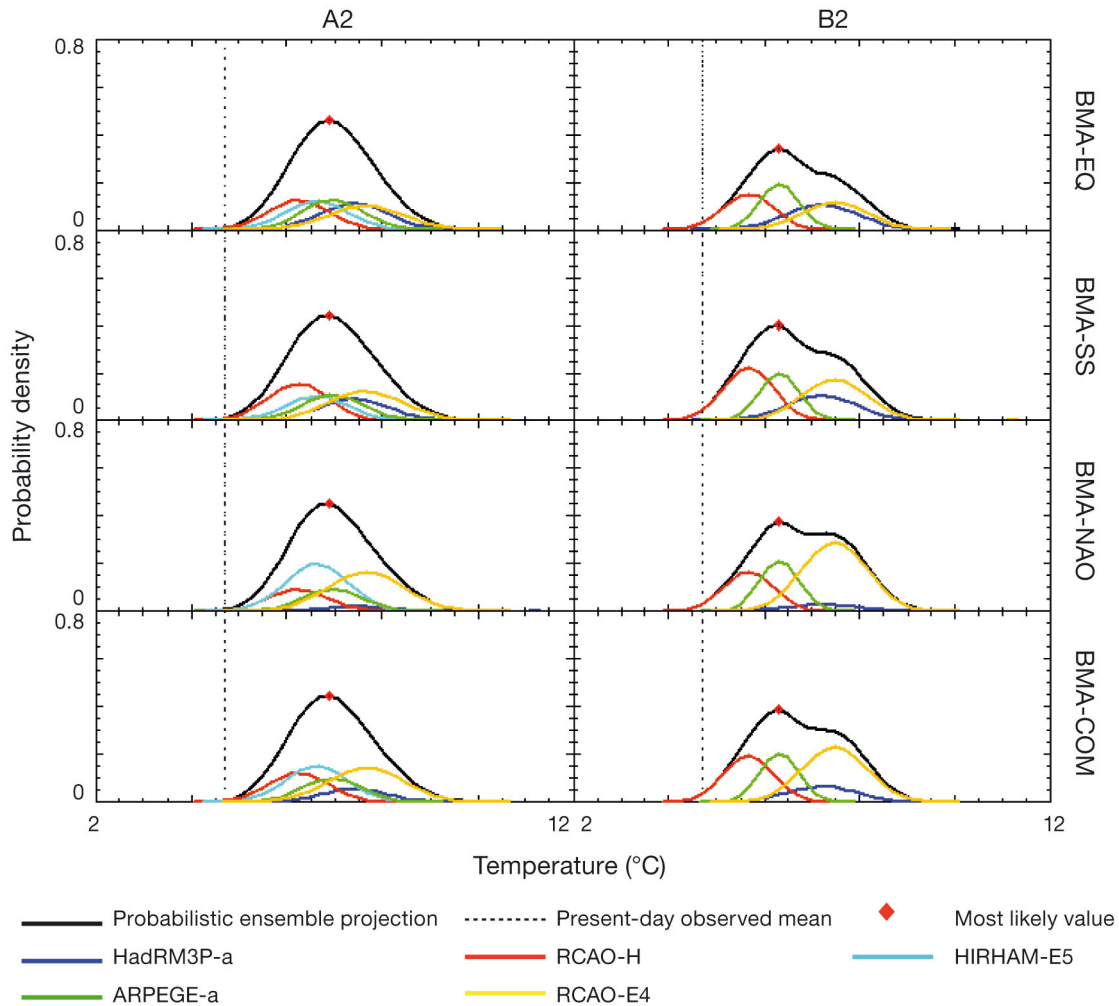

Fig. 8. Winter temperature projections for the A2 and B2 emissions scenarios

model. Choice of weighting had a large effect on the shape of the PDF. The 4 models formed 2 separate peaks. RCAO-E4 and HadRM3P-a both had most likely values of approximately 7.2 to 7.6°C, while RCAO-H and ARPEGE-a had most likely values of approximately 5.7 and 6.3°C, respectively. When these models were combined, the resulting distribution was bimodal, but the degree of bimodality depended on the weighting scheme applied. For example, when all models were weighted equally, the influence of RCAO-E4 was constrained. When weights were introduced, the high scores of RCAO-E4 in both spatiotemporal metrics and NAO representation made it a more influential model and the PDF became bimodal. For this parameter, season and emissions scenario, the shape of the ensemble PDF and the conclusions that might be drawn from it vary significantly when different weighting schemes are applied.

6.3. Summer temperature projections (2071–2100)

RCAO-E4 and RCAO-H were less skilful at simulating the summer spatial pattern than at simulating the winter spatial pattern and so had less influence on the calculation of summer projections. For the A2 scenario, the tails of the distribution PDF were long, ranging from 13 to 21°C, with a most likely projection of approximately 16.5°C regardless of the weighting system used (Fig. 9).

An interesting feature is that the heaviness of the upper tail varied depending on the weighting system used. Under BMA-SS weighting, the upper tail was thinner. Under BMA-NAO weighting, the influence of RCAO-E4 was greater, contributing to a heavier tail. This means that higher levels of probability are attached to the upper extremes of the ensemble PDF.

For the B2 scenario, there was a significant difference between the weights associated with RCAO-
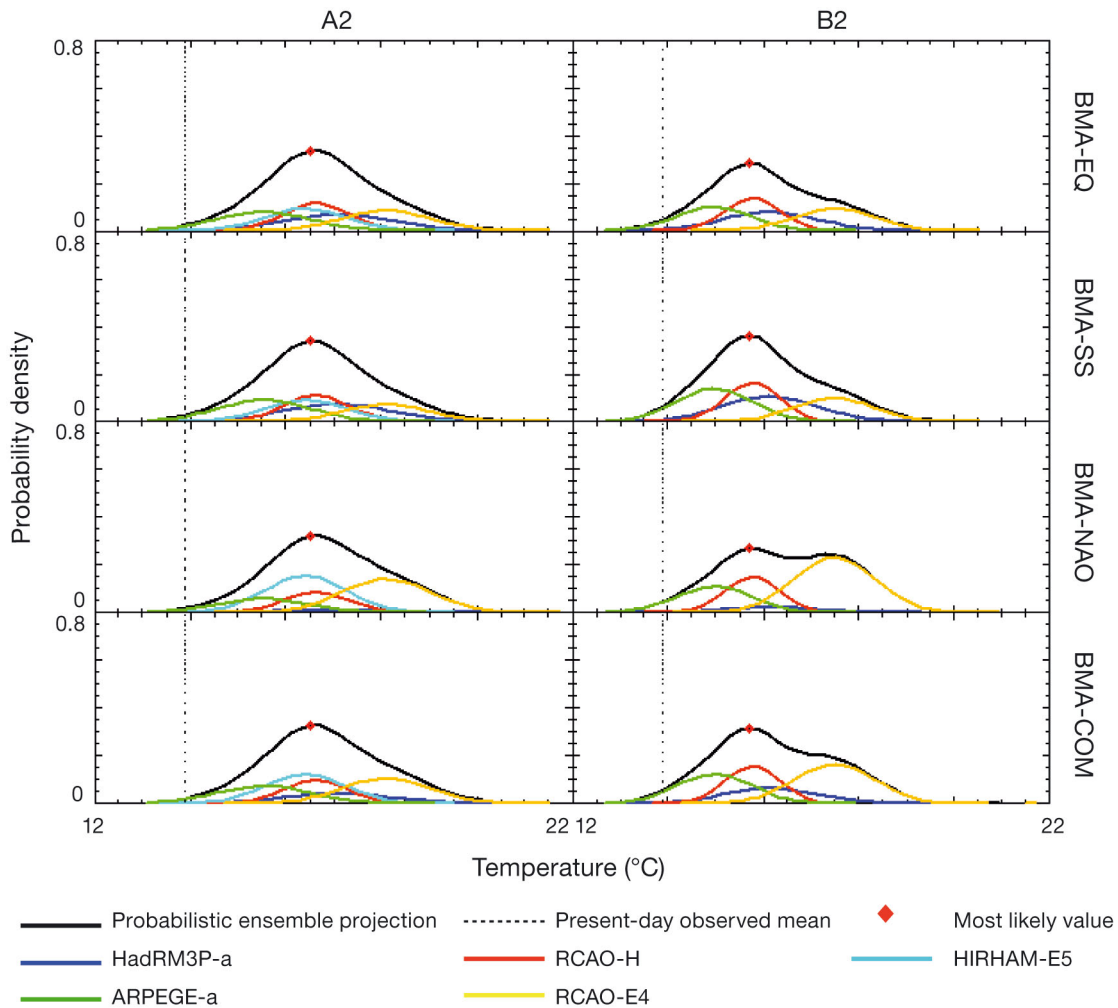

Fig. 9. Summer temperature projections for the A2 and B2 emissions scenarios
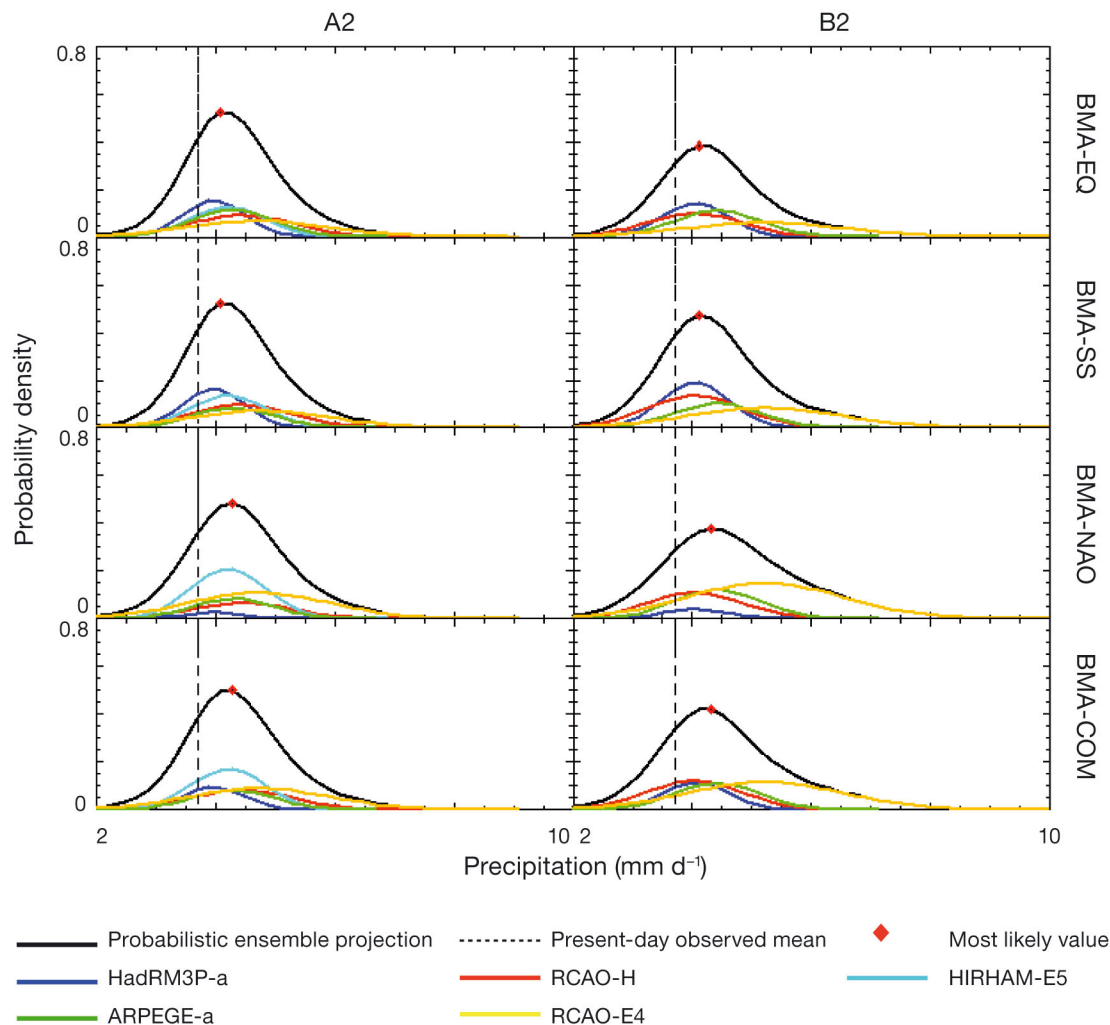
Fig. 10. Winter precipitation projections for the A2 and B2 emissions scenarios

E4 under the BMA-SS and BMA-NAO weighting schemes. As HIRHAM-E5 was not available with B2 forcing, RCAO-E4 became the most skilful model for representing the NAO. However, this model was not as skilful when it was scored based on spatiotemporal metrics.

The varying weights had significant effects on the probabilistic climate projections. When equal weights or skill-score-based weights were applied, the ensemble PDF had a heavy upper tail and a most likely value for future temperature of 15.7°C. However, when BMA-NAO or BMA-COM weighting was applied, the contribution of RCAO-E4 became much greater and the influence of other models less skilful at representing the NAO was constrained. As a result, the weighted ensemble PDF became bimodal, with peaks at 15.7 and 17.6°C. The bimodal PDF was most pronounced for BMA-NAO weighting and less pronounced using BMA-COM weighting.

It appears that under the B2 forcing scenario, the shape of the temperature ensemble PDF in both winter and summer was significantly influenced by the choice of weighting scheme. Therefore, it is vital that the weightings chosen are genuinely representative of the predictive skill of the model. The BMA-COM approach is clearly preferable, as more information about model skill in the present day is incorporated into this weighting scheme than the others.

6.4. Winter precipitation projections (2071–2100)

Fig. 10 shows the BMA-EQ, BMA-SS, BMA-NAO and BMA-COM ensemble projections for winter A2 and B2 precipitation. When equal weights or skill-score-based weights were applied, the ensemble PDF based on A2 winter precipitation data was normal. When BMA-NAO or BMA-COM weightings were

used, this distribution became slightly less peaked. In the first 2 projections, the HadRM3P-a projection dominated. The 2071–2100 output from this model had a smaller standard deviation compared to other models. This resulted in a likelihood distribution with a greater peak relative to other models, which had an impact on the shape of the ensemble PDF under the BMA-EQ and BMA-SS approaches. However, HadRM3P-a demonstrated low skill in the NAO assessment. As such, its weighting under the BMA-NAO and BMA-COM approaches was lower than that of the other models, and the lower weighting diminished its influence on the final projection.

In this instance, likelihood associated with the ensemble PDF was influenced by the projections of HadRM3P-a, yet the NAO assessment indicated that this model did not perform well in simulating the large-scale drivers of Irish climate. As such, its likelihood function was potentially over-confident. However, the advantage of the Bayesian approach is that further information about model skill, such as performance at simulating the NAO, can be incorporated into the projection to reflect more fully the state of knowledge about model skill.

For winter B2 precipitation, when equal weights or skill-score-based weights were applied, the ensemble PDF was quite similar, with a most likely value of between 4.0 and 4.2 mm d^{-1}. Again, HadRM3P-a had a notably peaked likelihood distribution, and, when combined with a skilful spatiotemporal skill score, the resultant ensemble PDF appeared more robust.

However, the addition of NAO skill information into the weighting changed the projection considerably. HadRM3P-a demonstrated a low level of skill in simulating the NAO; therefore, its lower weighting when this information was included dampened the initial high confidence associated with its projection, limiting its influence on the weighted ensemble PDF. RCAO-E4 became the key contributor to the ensemble PDF, as this model captured the NAO quite well. As such, the projection became heavier in the upper tail. The most likely projected precipitation value under BMA-NAO weighting was 4.3 mm d^{-1}. When both skill scores and NAO information were combined in the BMA-COM weightings, the most likely projected precipitation value was approximately 4.1 mm d^{-1}.

6.5. Summer precipitation projections (2071–2100)

Fig. 11 shows the BMA-EQ, BMA-SS, BMA-NAO and BMA-COM ensemble projections for summer A2 and B2 precipitation. For the A2 scenario, when equal weights or skill-score-based weights were applied, the ensemble PDF was heavier on the lower tail, suggesting that extreme low values of precipitation were more likely than extreme high values of precipitation under the A2 scenario. The influence of HadRM3P-a was evident here. This model tended towards drier conditions in the control period, though there was no significant systematic bias found or corrected for. However, it is important to note that bias in the control period may not be representative of bias under future forcing conditions. Model errors and biases may not remain constant under different forcing conditions, and, as such, the skill scores calculated for the control period may not reflect the skill of the model in a future time period. Incorporating other forms of skill assessment into the projection may add to the reliability of the projection.

The addition of NAO skill information into the weighting resulted in a more confident projection, illustrated by enhancement of the distribution peak. As HadRM3P-a demonstrated a low level of skill in simulating the NAO, its influence on the weighted ensemble PDF was dampened. The most likely projected precipitation value under BMA-NAO weighting was approximately 2.15 mm d^{-1}, a decrease of 0.3 mm d^{-1} compared with the control period. This was a 12% decrease, amounting to 9 mm less precipitation over the course of a month. However, it must be noted that, while this was the value with the highest likelihood, there was a range of both higher and lower values modelled by the RCMs, and the full range of these outcomes must be considered for the purposes of robust climate planning.

When the spatial r-score signifies that no strong co-variation occurs between present-day observed and modelled patterns, the bias patterns cannot be considered systematic and are not corrected for. While forming projections using the relative change within the model would possibly overcome such errors, such an approach makes the unverifiable assumption that the relative change within the model is the climate change signal and that errors will not fluctuate over time. Inevitably, any method used to develop projections of future climate is susceptible to uncertainties of different kinds, and communication of these uncertainties becomes the key issue.

For the B2 scenario, the overall shape of the weighted ensemble PDF varied considerably depending on the weighting used. When equal weights were applied, the ensemble PDF was heavy on the upper tail, but when skill scores were used, the distribution started to become bimodal. HadRM3P-a, which had a smaller standard deviation in the future than the other
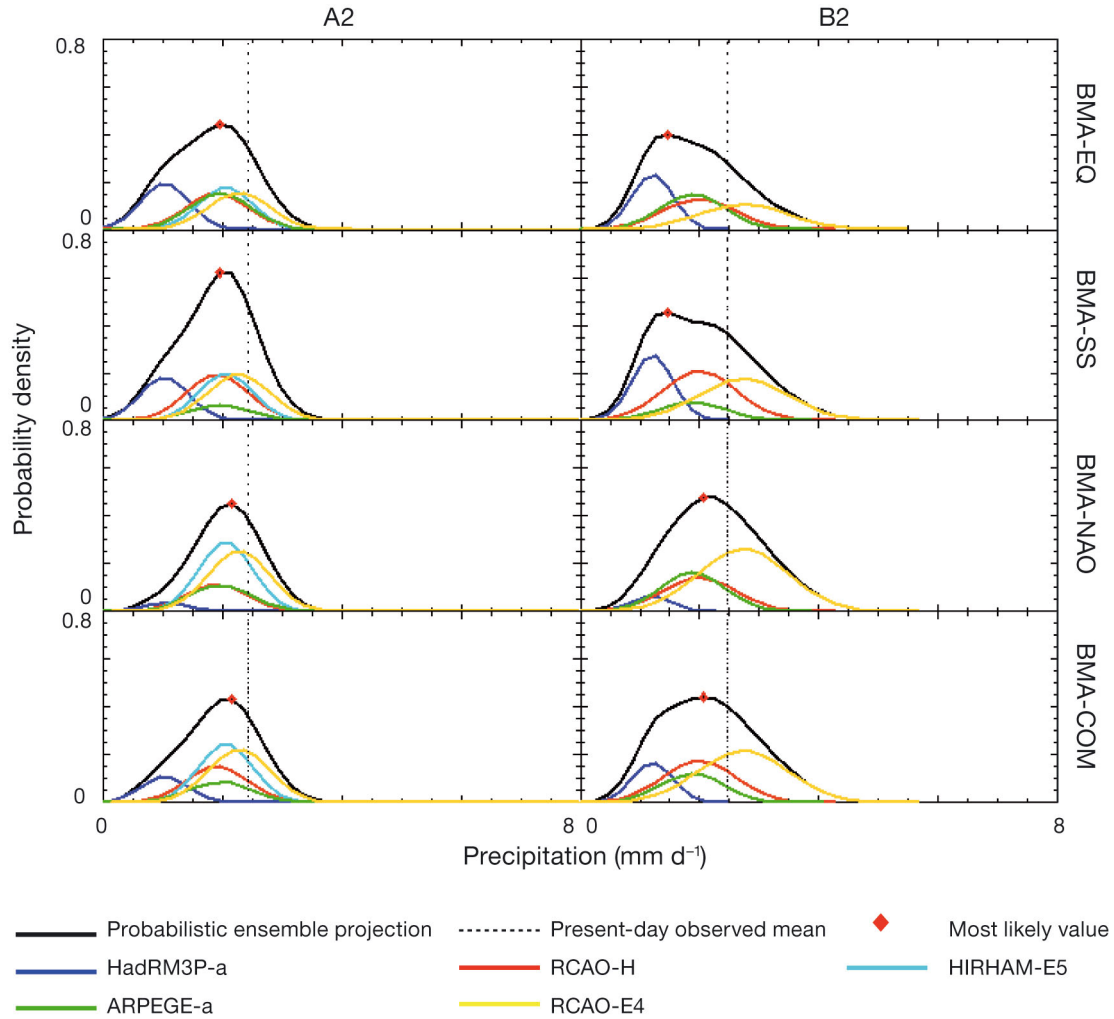
Fig. 11. Summer precipitation projections for the A2 and B2 emissions scenarios

models and simulated drier summer conditions in the control period that the other models, was a key contributor to this projection. The addition of NAO skill information into the weighting resulted in a distribution curve that was approximately normal. As RCAO-E4 modelled the NAO with considerable skill in the control period, it became the key contributor to the projection when NAO weights were used. As HadRM3P-a demonstrated a low level of skill in simulating the NAO, its influence on the weighted ensemble PDF was dampened.

In this instance, neither skill scores or NAO information alone offered an optimal approach for generating a projection, as the most influential models in both cases were outlying models with a degree of uncertainty attached to their projections. The combined weighting approach may offer enhanced reliability, and the Bayesian methodology offers the ability to quantify the uncertainty associated with these

divergent precipitation signals and combine them into a single PDF. The most likely projected precipitation value under BMA-COM weighting was approximately 2.07 mm d^{-1}, a decrease of 0.38 mm d^{-1} or 15% compared with the control period. Again, while this was the value with the highest likelihood associated with it under the BMA-COM weighting approach, there was a range of both higher and lower projections to be taken into account.

Table 3 summarizes the differences in the lower (5%) and upper (95%) distribution extremes as weighting schemes are varied. For most parameters, changes were small, but there was the potential for significant differences. For example, summer temperature at the 95% interval varied by up to half a degree under the B2 forcing scenario, which was a considerable difference due to weighting alone. Due to limits on model independence, the size of this ensemble was rather small, but, undoubtedly, a key

Table 3. Lower (5%) and upper (95%) distribution extremes as Bayesian model averaging (BMA) weighting schemes are varied

Weighting scheme	Emissions scenario	Winter		Summer	
		5%	95%	5%	95%
Temperature (°C)					
BMA-EQ	A2	5.18	8.88	14.32	19.32
	B2	4.89	8.59	13.91	18.51
BMA-SS	A2	5.18	8.88	14.32	19.12
	B2	4.79	8.69	13.71	18.51
BMA-NAO	A2	5.18	8.98	14.52	19.52
	B2	4.89	8.79	13.91	19.01
BMA-COM	A2	5.18	8.88	14.32	19.32
	B2	4.79	8.69	13.81	18.81
Precipitation (mm d^{-1})					
BMA-EQ	A2	2.70	6.20	0.37	3.27
	B2	2.71	6.31	0.54	3.74
BMA-SS	A2	2.70	6.20	0.37	3.27
	B2	2.61	6.51	0.49	3.99
BMA-NAO	A2	2.70	6.40	0.72	3.32
	B2	2.61	7.11	0.64	4.14
BMA-COM	A2	2.70	6.30	0.52	3.32
	B2	2.61	6.91	0.54	4.09

research question remaining is whether increasing ensemble size, while ensuring model independence, would reduce or increase this variability.

7. DISCUSSION

Results illustrate that mean-based skill assessments may not be a good indicator of a model's ability to simulate the dynamics of the climate system. However, it is important to note that mean temperature and precipitation values in a RCM are affected by many factors within the model and it is expected that multiple sources will account for the errors.

It is clear from these results that RCMs can respond quite differently to GCM deficiencies. In an analysis of mean circulation indices in GCMs, van Ulden et al. (2007) found positive westerly biases in the HadAM3H AGCM, ARPEGE (included as a variable resolution AGCM) and the ECHAM4-OPYC AOGCM, which also suggests that RCM errors in NAO representation arise from the boundary conditions supplied by the GCM drivers. However, only in ARPEGE-a is the westerly bias reported by van Ulden et al. (2007) evident in the regional simulation. The HadAM3H GCM, which is used to model many of the simulations in the PRUDENCE ensemble, has a tendency to model a steeper pressure gradient than that observed (Jacob

et al. 2007), which in turn would impact how the RCMs driven by this model, such as RCAO-H driven by HadAM3H, represent temperature and precipitation. These outcomes illustrate that more attention must be given to GCM skill when coupling GCMs and RCMs, as biases in the GCM can impact RCM performance and failure of the GCM to model the large-scale climate cannot be improved using an RCM.

To assess what impact the inclusion or exclusion of information about model performance has on the resulting projection, ensemble projections were generated using the case study models, so that the information gained about their representation of the NAO could be incorporated using the BMA technique.

Under the A2 emissions scenario, the mean ensemble PDFs for both winter and summer temperature changed little when different weighting schemes were used. However, the relative contributions of the ensemble members underlying that mean projection did vary. An important finding is that under the B2 scenario, the mean ensemble PDFs changed considerably for both winter and summer temperature, when different weighting schemes were applied. In summer especially, the BMA-SS approach of weighting based on aspects of performance in the control period resulted in an approximately normal curve with a single peak. However, when weights were employed that reflect the ability to capture the effects of the NAO in the control period (BMA-NAO), the resulting ensemble PDF was bimodal.

The mean projection for both winter and summer precipitation under the B2 scenario also changed considerably when different weighting schemes were used. Under the A2 emissions scenario, varying the weighting scheme impacted the overall likelihood and the shape of the lower tail of the summer precipitation distribution. As such, there was uncertainty over the likelihood attached to extremely low levels of precipitation in summer. Although these values were not the most probable, having much lower likelihood attached to them than the most likely projection, such low levels of summer precipitation would have considerable impacts associated with them if they were to occur. Therefore, it is in the interest of robust climate planning to take these values into consideration.

Under the B2 emissions scenario the ensemble precipitation PDF varied with respect to both shape and maximum likelihood when different weightings were used. Under BMA-NAO and BMA-COM weighting, the distributions were approximately normal, while, under BMA-EQ and BMA-SS weighting, the distributions were skewed to the left. Projections for the

B2 forcing scenario had a greater degree of uncertainty attached to them, particularly with regards to the upper extremes of the precipitation distribution, as the choice of weighting scheme introduced another layer of variability into the climate modelling process.

8. CONCLUSIONS

It is important to note that the models investigated in this research were examined with respect to a specific domain and assessed on their ability to model particular aspects of the climate system. As such, the skill metrics themselves are applicable only to this domain and study. For other domains and climate parameters, it is quite likely that the models would perform differently, exhibiting different levels of skill. In any research involving climate models, the model must be assessed to ensure that simulations are credible for the domain and parameters of interest.

However, a more significant conclusion to be drawn from these findings is that there is a need to move towards more comprehensive weighting approaches that incorporate information about model skill in areas beyond the mean climate state. NAO skill assessment has demonstrated that skill-score-based assessments of the mean climate state may not identify deficiencies in the simulation of large-scale climate drivers. As such, formulating future climate projections based on skill scoring assessments was found to be an unreliable approach, and techniques which incorporate a fuller picture of model skill into the formulation of projections are required.

The effect of variation in the weighting scheme was not always visible in the mean ensemble PDF, which highlights another important point. Even when the choice of weighting scheme does not change the shape of the ensemble PDF considerably, the data underlying the projection may be altered, as different models become the key contributors to the projections. When such uncertainties are associated with the underlying data, it is highly important that these uncertainties are characterized and communicated, even if the mean projection is not largely affected. For example, the confidence that might be attached to an ensemble PDF which is dominated by a single RCM projection would be very different to the confidence that might be attached to an ensemble PDF in which all RCMs converge and contribute equally, yet the overall shape of the ensemble PDF could be very similar.

The ensemble used in this study was limited to 4 to 5 models to satisfy the requirement for model independence. As such, the posteriors may have been significantly affected by a single model. Extending the ensemble size, while maintaining model independence, would increase the robustness of results.

As such it is important, when providing information for decision-makers, to consider not only the end product, the climate projection, but also the techniques used to produce that projection and the uncertainties associated with them. Yet it is worth noting that no technique can produce projections with absolute confidence. While skill at simulating the present is an indicator of model reliability, it is not a guarantee of accurate future projections; therefore, the possibility exists that the actual future climate will fall outside of the modelled range. However, while model outputs may always have an element of uncertainty associated with them, the insights into and greater understanding of the climate system that we gain by modelling it will have immense value (Smith & Stern 2011).

LITERATURE CITED

Abramowitz G (2010) Model independence in multi-model ensemble prediction. Aust Meteorol Oceanogr J 59:3–6

Bouwer L, Bubeck P, Aerts JC (2010) Changes in future flood risk due to climate and development in a Dutch polder area. Glob Environ Change 20:463–471

British Irish Council (2003) Scenarios of climate change for islands within the BIC Region. Met Office Hadley Centre for Climate Prediction and Research, Exeter

Brown JD (2004) Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Trans Inst Br Geogr 29:367–381

Castro M, Fernandez C, Gaertner MA (1993) Description of a mesoscale atmospheric numerical model. In Diaz JI, Lions JL (eds) Mathematics, climate and environment. Masson, Paris, p 230–253

Christensen JH, Bøssing Christensen O, Lopez P, van Meij-gaard E, Botzet M (1996) The HIRHAM4 regional atmospheric climate model. Scientific Report 96-4, Danish Meteorological Institute

Christensen JH, Kjellström E, Giorgi F, Lenderink G, Rummukainen M (2010) Weight assignment in regional climate models. Clim Res 44:179–194

Coppola E, Giorgi F, Rauscher SA, Piani C (2010) Model weighting based on mesoscale structures in precipitation and temperature in an ensemble of regional climate

models. Clim Res 44:121–134

Déqué M, Marquet P, Jones RG (1998) Simulation of climate change over Europe using a global variable resolution general circulation model. Clim Dyn 14:173–189

Doms G, Schattler U (2002) A description of the nonhydrostatic regional model LM. I. Dynamics and numerics. Consortium for Small-Scale Modeling Report, Deutscher Wetterdienst, Offenbach

Döscher R, Willen U, Jones C, Rutgersson A, Meier HEM, Hansson U, Graham LP (2002) The development of the regional coupled ocean-atmosphere model RCAO. Boreal Environ Res 7:183–192

Gates WL, Boyle JS, Covey C, Dease CG and others (1999) An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). Bull Am Meteorol Soc 80:29–55

Giorgi F, Marinucci MR, Bates GT (1993) Development of a 2nd generation regional climate model (REGCM2). 1. Boundary-layer and radiative-transfer processes. Mon Weath Rev 121:2794–2813

Giorgi F, Mearns LO (2003) Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. Geophys Res Lett 30:1629–1633

Goodess CM, Palutikof JP (1998) Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. Int J Climatol 18:1051–1083

Grimit EP, Mass CF (2002) Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. Weather Forecast 17:192–205

Harris GR, Collins M, Sexton DMH, Murphy JM, Booth BBB (2010) Probabilistic projections for 21st century European climate. Nat Hazards Earth Syst Sci 10:2009–2020

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14:382–401

Jacob D, Podzun R (1997) Sensitivity studies with the regional climate model REMO. Meteorol Atmos Phys 63:119–129

Jacob D, Barring L, Christensen OB, Christensen JH and others (2007) An inter-comparison of regional climate models for Europe: model performance in present-day climate. Clim Change 81:31–52

Jenkinson AF, Collison FP (1977) An initial climatology of gales over the North Sea. Synoptic climatology branch memorandum, Book 62. Meteorological Office, Bracknell

Jones PD, Hulme M, Briffa KR (1993) A comparison of Lamb circulation types with an objective classification scheme. Int J Climatol 13:655–663

Kiely G (1999) Climate change in Ireland from precipitation and streamflow observations. Adv Water Resour 23: 141–151

Linderson ML (2001) Objective classification of atmospheric circulation over southern Scandinavia. Int J Climatol 21: 155–169

Lopez A, Tebaldi C, New M, Stainforth D, Allen M, Kettleborough J (2006) Two approaches to quantifying uncertainty in global temperature changes. J Clim 19:4785–4796

Lucarini V, Speranza A, Vitolo R (2007) Parametric smoothness and self-scaling of the statistical properties of a minimal climate model: What beyond the mean field theories? Physica D 234:105–123

Moberg A, Jones PD (2004) Regional climate model simula-

tions of daily maximum and minimum near-surface temperatures across Europe compared with observed station data 1961–1990, Clim Dyn 23:695–715

New M, Lopez A, Dessai S, Wilby R (2007) Challenges in using probabilistic climate change information for impact assessments: an example from the water sector. Philos Trans R Soc Lond A 365:2117–2131

Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. Clim Dyn 16:123–146

Raisanen J, Hansson U, Ullerstig A, Doscher R and others (2004) European climate in the late twenty-first century: regional simulations with two driving global models and two forcing scenarios. Clim Dyn 22:13–31

Roeckner E, Arpe K, Bengtsson L, Christoph M and others (1996) The atmospheric general circulation model ECHAM-4: model description and simulation of present day climate. Max Planck Institute for Meteorology, Hamburg

Rinke A, Dethloff K, Cassano JJ, Christensen JH and others (2006) Evaluation of an ensemble of Arctic regional climate models: spatiotemporal fields during the SHEBA year. Clim Dyn 26:459–472

Sanchez E, Romera R, Gaertner MA, Gallardo C, Castro M (2009) A weighting proposal for an ensemble of regional climate models over Europe driven by 1961–2000 ERA40 based on monthly precipitation probability density functions. Atmos Sci Lett 10:241–248

Schmittner A, Latif M, Schneider B (2005) Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. Geophys Res Lett 32:L23710, doi:10.1029/2005GL024368

Smith LA, Stern N (2011) Uncertainty in science and its role in climate policy. Philos Trans R Soc Lond A 369:1–24

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos Trans R Soc Lond A 365:2053–2075

Tebaldi C, Mearns LO, Nychka D, Smith RL (2004) Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. Geophys Res Lett 31:L24213, doi:10.1029/2004GL021276

Tracton MS, Kalnay E (1993) Operational ensemble prediction at the National Meteorological Center—practical aspects. Weather Forecast 8:379–398

van Ulden A, Lenderink G, van den Hurk B, van Meijgaard E (2007) Circulation statistics and climate change in Central Europe: PRUDENCE simulations and observations. Clim Change 81:179–192

Vidale PL, Lüthi D, Frei C, Seneviratne SI, Schär C (2003) Predictability and uncertainty in a regional climate model. J Geophys Res 108(D18):23, doi:10.1029/2002JD 002810

Whitaker JS, Loughe AF (1998) The relationship between ensemble spread and ensemble mean skill. Mon Weather Rev 126:3292–3302

Wilby RL, O'Hare G, Barnsley N (1997) The North Atlantic Oscillation and British Isles climate variability, 1865–1996. Weather 52:266–276

Yun WT, Stefanova L, Mitra AK, Kumar T, Dewar W, Krishnamurti TN (2005) A multi-model superensemble algorithm for seasonal climate prediction using DEMETER forecasts. Tellus A 57:280–289