# Multimodel Future Projections of Wintertime North Atlantic and North Pacific Tropospheric Jets: A Bayesian Analysis

D. WHITTLESTON

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts*

K. A. MCCOLL

*Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

D. ENTEKHABI

*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Manuscript received 22 May 2017, in final form 13 December 2017)

ABSTRACT

The impact of future greenhouse gas forcing on the North Atlantic and North Pacific tropospheric jets remains uncertain. Opposing changes in the latitudinal temperature gradient—forced by amplified lower-atmospheric Arctic warming versus upper-atmospheric tropical warming—make robust predictions a challenge. Despite some models simulating more realistic jets than others, it remains the prevailing approach to treat each model as equally probable (i.e., democratic weighting). This study compares democratically weighted projections to an alternative Bayesian-weighting method based on the ability of models to simulate historical wintertime jet climatology. The novel Bayesian technique is developed to be broadly applicable to high-dimensional fields. Results show the Bayesian weighting can reduce systematic bias and suggest the wintertime jet response to greenhouse gas forcing is largely independent of this historical bias (i.e., not state dependent). A future strengthening and narrowing is seen in both winter jets, particularly at the upper levels. The widely reported poleward shift at the level of the eddy-driven jet does not appear statistically robust, particularly over the North Atlantic, indicating sensitivity to current model deficiencies.

## 1. Introduction

The tropospheric westerly jet streams play a central role in much of the weather and climate variability in the wintertime extratropics. In the Southern Hemisphere a single jet circumvents the globe, whereas in the Northern Hemisphere (NH) two distinct jets form over the Atlantic and Pacific (Fig. 1a). These jets act as waveguides for regions of preferential cyclone activity, or storm tracks (Hoskins and Ambrizzi 1993; Lorenz and Hartmann 2003), which are responsible for wind, precipitation, and temperature extremes (see review by Shaw et al. 2016). In fact, the north–south wobbling of these jets represents the primary modes of atmospheric variability away from the tropics (Wittman et al. 2005).

Two dynamical processes are responsible for the formation of these jets. The first is poleward-moving air accelerating westward to create a jet on the edge of the Hadley cell. The second is eddy heat flux and momentum flux convergence caused by atmospheric waves forming in regions of enhanced baroclinicity. These two jets are often separated conceptually and referred to as the subtropical and eddy-driven jets, respectively (Hartmann 2007). In reality, both processes often operate and interact continuously making it difficult to attribute a given jet to either process (Lee and Kim 2003).

Predicting future changes in the jets, and corresponding shifts in circulation, would provide valuable

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-17-0316.s1.

*Corresponding author*: David Whittleston, dwhittle@mit.edu

DOI: 10.1175/JCLI-D-17-0316.1

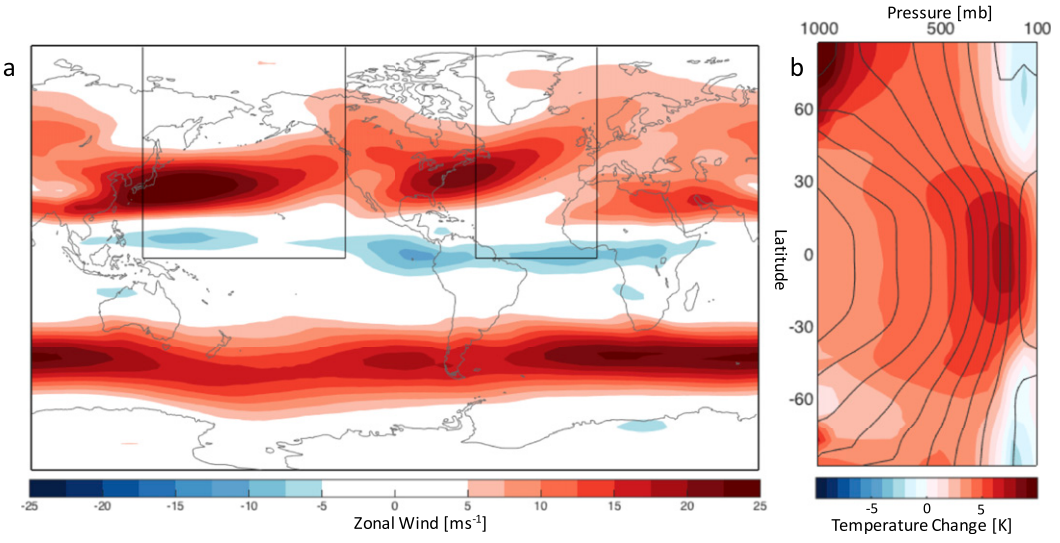


FIG. 1. (a) Climatology of DJF zonal wind at 500 hPa (1950–2005; NCEP–NCAR reanalysis). Boxes show North Pacific (0°–90°N, 135°W–125°E) and North Atlantic (0°–90°N, 0°–60°W) domains. (b) Latitude–height cross section showing the projected DJF temperature change (shading) at the end of the twenty-first century (RCP8.5; 2075–99) relative to historical simulations (1950–2005). Contours show the DJF temperature climatology in historical CMIP5 simulations at intervals of 10 K from 290 K at left to 210 K at right. Values were calculated using the AEM of all CMIP5 simulations in Table 1.

information to assess the societal impacts of climate change. Any modification of global temperature distributions may impact circulation through changes to both temperature gradients and stability (e.g., Butler et al. 2010). Unlike the thermodynamics response, however, the dynamic response to increase greenhouse gas forcing is not well understood (Shepherd 2014).

It has been hypothesized that Arctic amplification—the accelerated warming of the Arctic relative to the global mean—will decrease the latitudinal temperature gradient, slow down the zonal jets, and cause them to shift equatorward (Francis and Vavrus 2012; Liu et al. 2012). While Deser et al. (2004) and others have shown in targeted modeling experiments that a warming Arctic can induce an equatorward shift of the midlatitude jets, this does not mean that a future Arctic amplification will necessitate such a shift (Barnes and Screen 2015). In fact, many climate projections simulate a poleward or negligible shift of the tropospheric jets by the end of the twenty-first century, despite all exhibiting Arctic amplification (Barnes and Polvani 2015).

This apparent contradiction can be reconciled by model predictions of enhanced warming in the tropical upper troposphere resulting from changes in moist adiabatic lapse rate (Vallis et al. 2015). That is, while the lower-level equator–pole temperature gradient is expected to decrease, the upper-level equator–pole temperature gradient is expected to increase (Fig. 1b; IPCC 2007, their Fig. 10.7). Which of these will emerge as the dominant forcing on future jets and storm tracks remains an open question.

The complexity of midlatitude dynamics makes it challenging to predict how jets will respond to climate change using simple theoretical reasoning alone. Atmosphere–ocean general circulation models (AOGCMs) attempt to realistically simulate the global climate system and its response to greenhouse gas emissions. These comprehensive models aspire to be a culmination of the scientific community's current knowledge and computational ability. Coordinated modeling experiments conducted by a range of institutions, such as phase 5 of the Coupled Model Intercomparison Project (CMIP5), attempt to quantify the uncertainty in projections that results from our incomplete understanding of nature and its imperfect representation in models (Knutti et al. 2017). Vallis et al. (2015) studied the CMIP5 archive to investigate future changes in large-scale circulation, including the midlatitude jets. As is the prevailing approach, their study assumes ''model democracy''; each of the CMIP5 models is thought to be equally skillful and its projection equally likely (Knutti 2010). However, there is clear evidence that some models are more capable of modeling particular phenomena compared to others (Gleckler et al. 2008).

An alternative to model democracy is to weigh model projections based on its skill to realistically simulate historical conditions. The key assumption in this approach is that if a model faithfully reproduces an observed phenomenon, the key processes responsible for that phenomenon are being realistically simulated. It follows that the model is more likely to respond realistically to

TABLE 1. List of CMIP5 models, including number of ensemble runs, used in this study. Square brackets denote the same model run with different physics packages. Ensembles refer to realizations of the same model using different but equally realistic initial conditions. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

| Model name | Model center | No. of ensembles |
|---|---|---|
| BCC_CSM1.1 | Beijing Climate Center, China Meteorological Administration | 1 |
| BCC_CSM1.1(m) | Beijing Climate Center, China Meteorological Administration | 1 |
| CCSM4 | National Center for Atmospheric Research | 5 |
| CESM1(CAM5) | National Center for Atmospheric Research | 2 |
| CNRM-CM5 | Centre National de Recherches Météorologiques | 1 |
| GFDL CM3 | Geophysical Fluid Dynamics Laboratory | 1 |
| GISS-E2-R [P1] | NASA Goddard Institute for Space Studies | 2 |
| GISS-E2-R [P3] | NASA Goddard Institute for Space Studies | 2 |
| INM-CM4.0 | Institute of Numerical Mathematics | 1 |
| IPSL-CM5A-LR | L'Intitut Pierre-Simon Laplace | 4 |
| IPSL-CM5A-MR | L'Intitut Pierre-Simon Laplace | 1 |
| IPSL-CM5B-LR | L'Intitut Pierre-Simon Laplace | 1 |
| MIROC5 | Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies | 2 |
| MIROC-ESM | Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies | 1 |
| MPI-ESM-LR | Max Planck Institute for Meteorology | 3 |
| MPI-ESM-MR | Max Planck Institute for Meteorology | 1 |
| NorESM1-M | Norwegian Climate Centre | 1 |

future forcing. This holds providing future mean climate is within the envelope of historical variability (i.e., the climate system has drifted but not converted to another mode of variability). Knutti et al. (2017) provide a concise overview of the arguments for and against weighing models. The main challenge in pursuing this approach is how to convert a "model performance metric" (i.e., how well a model simulates observations) into weights in an objective manner.

The two goals of this study were to 1) analyze the future changes in the NH wintertime atmospheric jets as simulated by the CMIP5 ensemble by 2) developing and applying a novel Bayesian-weighting scheme. The approach relies on decomposition of the simulated atmospheric fields onto dominant modes of variability. Section 2 describes the data and methods used in our study. The historical bias associated with the democratic weighting of the jets is investigated in section 3. In section 4 we outline the Bayesian-weighting approach and verify its use on historical data. We then use these weights to investigate changes in the jets at the end of the twenty-first century in section 5. Section 6 offers some conclusions.

## 2. Data and methods

In this study we compare simulations performed in CMIP5 (Taylor et al. 2012) to our "best guess" of historical conditions—provided by atmospheric reanalysis data. Atmospheric zonal wind and temperature data spanning from 1000 to 100 hPa were from the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996). Although a number of new reanalysis products that offer improved quality are now available, none match the length of the NCEP–NCAR observational record. Long-duration records are required for the robust calculation of eigenvalues outlined in the methodology for our Bayesian-weighting approach (section 4). We analyze the output from 30 CMIP5 simulations from 17 models (Table 1). Our approach considers all simulations independently (i.e., we do not attempt to consider the dependence of simulations from the same model). This assumption is not strictly true as ensembles of the same model are clearly related and many models from different groups exchange both ideas and code. To address such dependence Knutti et al. (2017) suggest an approach that penalizes similar models. However, this has the inherent limitation of reducing model weights if they converge on reality. The CMIP5 data were obtained through the Earth System Grid Federation data portal (http://pcmdi9.llnl.gov/) and represent the complete set available at the time of access. The simulations were performed at a range of resolutions so were first mapped onto the common 2.5° NCEP–NCAR reanalysis grid using triangulation-based cubic interpolation.

Our analysis focuses on December–February zonal wind $u$ calculated from monthly mean data. Winter is when the northward shift of the NH jets is least consistent

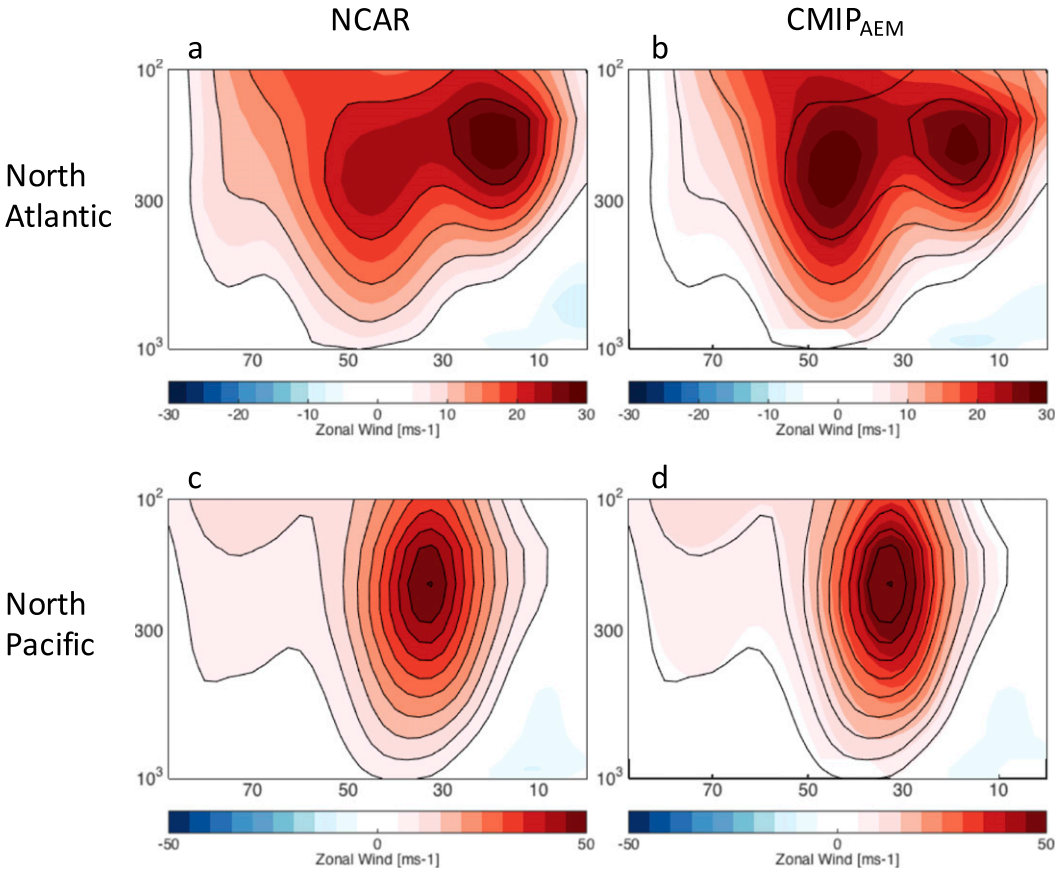


FIG. 2. Climatology (1950–2005) of zonally averaged DJF zonal wind over the North Atlantic in (a) NCEP–NCAR reanalysis and (b) the CMIP$_{AEM}$. CMIP$_{AEM}$ refers to the AEM of the CMIP5 simulations. (c),(d) As in (a),(b), but for the North Pacific domain. Contours show the reanalysis climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$.

from year to year and model to model, and both the upper- and lower-tropospheric temperature gradients are thought to play a significant role (Harvey et al. 2014; Simpson et al. 2014). To characterize the two jets we zonally average atmospheric data across the North Pacific (135°E–125°W) and North Atlantic (60°W–0°) domains shown in Fig. 1. These regions were chosen to be consistent with earlier analyses (Eichelberger and Hartmann 2007; Barnes and Polvani 2013). We analyze latitude–height cross sections, as opposed to geometric indices (e.g., Woollings et al. 2010), so we can assess the structures of the jets throughout the depth of the troposphere and to allow for visual verification of our Bayesian-weighting method.

Our historical period of analysis was constrained to the NCEP–NCAR data period spanning January 1950–December 2005 (55 DJF winters). Historical CMIP5 simulations were forced by observed greenhouse gas emissions, anthropogenic aerosol, and solar and volcanic activities and so represent the scientific community's best attempt to model the historical jets. CMIP5 projections of future climate are from the representative concentration pathway 8.5 (RCP8.5) emission scenario in which greenhouse gas emissions result in top-of-atmosphere radiative forcing of 8.5 W m$^{-2}$ by 2100 (van Vuuren et al. 2011). RCP8.5 represents the most extreme CMIP5 future emissions scenario and was chosen to maximize the signal-to-noise ratio in our analysis. We define the future state of the jets as their climatology during the last 25 years of the twenty-first century (2075–99).

## 3. Historical jets in CMIP5

The contrasting structure of the historical North Atlantic and North Pacific jets is evident in Fig. 2. Weak Hadley circulation and a zone of baroclinicity that is relatively far north result in distinct subtropical and eddy-driven jets over the North Atlantic ("separated jets"; Fig. 2a). While the subtropical jet is confined to the upper levels, the eddy-driven jet extends through the depth of the troposphere. Over the North Pacific, strong

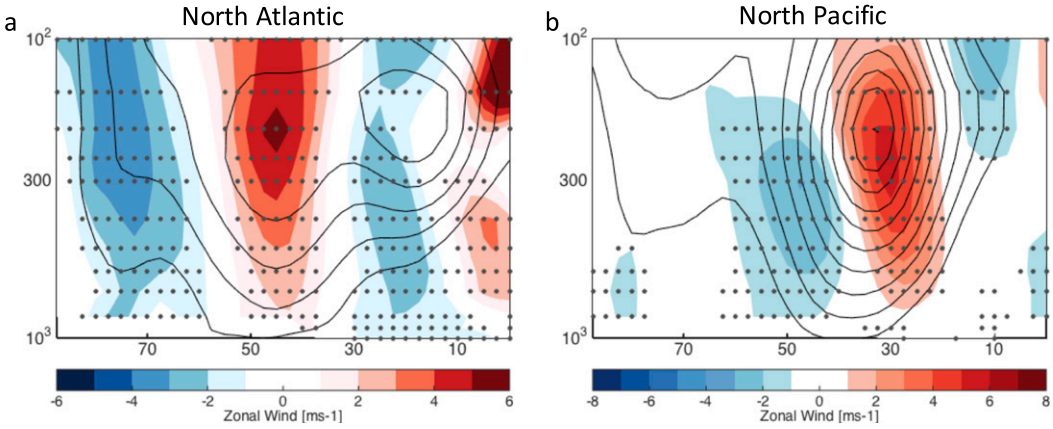


FIG. 3. Difference between NCEP–NCAR reanalysis and CMIP5 AEM zonal wind climatology (1950–2005) for the (a) North Atlantic and (b) North Pacific. Contours show the reanalysis climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$. Stippling marks differences significant at $\alpha_{FDR} = 0.1$ on a two-tailed test (see the appendix).

Hadley circulation and an equatorward zone of baroclinicity result in a single jet maxima (''combined jets''; Fig. 2c) (Li and Wettstein 2012). Figures 2b,d show the arithmetic ensemble mean (AEM) of the CMIP5 model jets. They successfully simulate the general differences between the zonal wind climatology in two regions, exhibiting clearly distinct jets in the Atlantic and combined jets in the Pacific.

While the broad features of the different jets are distinguishable there are clear differences between the AEM of the CMIP simulations and reanalysis climatology (Figs. 2b,d and 3). These broad differences are independent of the reanalysis product (Fig. S1 in the supplemental material). Over the North Pacific, we see a clear bias in the latitudinal position of the North Pacific jet, particularly in the lower troposphere. This is consistent with analysis by Barnes and Polvani (2013) that focused on the eddy-driven jet. This latitudinal shift also corresponds to a positive bias in the jet strength in the upper troposphere at its maximum (Figs. 2d and 3b).

In the North Atlantic there is a weak bias in the southern core and significant strong bias in the northern core. The strong bias at the latitude of the eddy-driven jet maxima is consistent with findings of positive bias in DJF jet strength over the North Atlantic in the CMIP3 ensemble (Woollings and Blackburn 2012). This bias is surprising given the large range of eddy-driven jet latitudes observed across the different models (Barnes and Polvani 2013), which one might expect to appear in the ensemble average as a weakening and spreading of the jet core. What we see in Fig. 3 is the opposite—a bias toward a stronger and narrower eddy-driven jet in the zonal mean. Much of this bias may be explained by the tendency of the North Atlantic storm tracks to be too zonal in models (Zappa et al. 2013) because jets that are more zonal would appear to have stronger and narrower cores in the zonal average. That is, the latitudinal tilt of the North Atlantic jet results in a ''smearing'' of the jet structure in the zonal average (Fig. 2a). Alternatively, this bias could be explained by unrealistically low variability in jet latitude over the North Atlantic (Gong et al. 2016).

## 4. Bayesian weighting

### a. Deriving Bayesian weights

The CMIP5 models simulate a range of distinct DJF zonal wind climatologies (Figs. S8–S11 in the supplemental material) that are democratically weighted to produce the AEM climatology shown in Figs. 2 and 3. From visual inspection it is clear some models simulate more unrealistic climatology than others—resulting in the bias shown in Fig. 3—so it would be reasonable to argue these models should be given less weight in predicting future conditions. To pursue this approach we must (i) choose a model quality metric and (ii) convert this metric into a weight for each model. As we wish to understand changes in the jets, the ability of a model to simulate historical zonal wind climatology (Fig. 2a) seems a reasonable choice for a model quality metric. This assumes that future mean changes will remain within the envelope of historical variability. While this could hypothetically not be the case, the authors are not aware of any literature suggesting either the North Atlantic or North Pacific mean jet will undergo such dramatic changes. Our approach also assumes that the dominant processes setting the mean state are also those that set the response to forcing.

Our goal is to assign low weight to CMIP model $i$ $\mathbf{m}_i$ if its simulation of historical zonal wind climatology agrees

poorly with reanalysis data **d**. To achieve approach (ii) we propose a method that builds on the Bayesian model averaging (BMA) framework developed by Min et al. (2004, 2007). We extend their approach so that high-dimensional observations **d**—such as zonal wind climatology—can be used as a metric for model quality. Starting from Bayes's rule we have

$$P(\mathbf{m}_i \,|\, \mathbf{d}) = \frac{P(\mathbf{d} \,|\, \mathbf{m}_i)P(\mathbf{m}_i)}{\sum_{j=1}^{N} P(\mathbf{d} \,|\, \mathbf{m}_j)P(\mathbf{m}_j)}, \tag{1}$$

where $P(\mathbf{m}_i \,|\, \mathbf{d})$ is the posterior, $P(\mathbf{d} \,|\, \mathbf{m}_i)$ is the likelihood, $P(\mathbf{m}_i)$ is the prior, and $\sum_{j=1}^{N} P(\mathbf{d} \,|\, \mathbf{m}_j)P(\mathbf{m}_j)$ is the marginal distribution. If we assume uniform priors, $P(\mathbf{m}_i) = 1/N$, and (1) simplifies to

$$P(\mathbf{m}_i \,|\, \mathbf{d}) = \frac{P(\mathbf{d} \,|\, \mathbf{m}_i)}{\sum_{j=1}^{N} P(\mathbf{d} \,|\, \mathbf{m}_j)}. \tag{2}$$

However, if there is a source of information on prior probabilities, this framework can incorporate them into the model weighting. With uniform priors, the weighting scheme is also equivalent to the maximum likelihood estimator. The ratio $P(\mathbf{m}_i \,|\, \mathbf{d})/P(\mathbf{d} \,|\, \mathbf{m}_j)$ corresponds to the Bayes factor often used for model selection (Gelman et al. 2003). Assuming the model realizations and observations follow a multivariate Gaussian distribution, $P(\mathbf{d} \,|\, \mathbf{m}_i)$ can be formulated as follows:

$$P(\mathbf{d} \,|\, \mathbf{m}_i) = \frac{1}{\sqrt{2\pi^q \det \boldsymbol{\Sigma}_d}} \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{m}_i)^{\mathrm{T}}(\boldsymbol{\Sigma}_d + \boldsymbol{\Sigma}_m)^{-1}(\mathbf{d} - \mathbf{m}_i)\right]. \tag{3}$$

Here the exponent represents the Mahalanobis distance between the observational data and model simulations. Here $q$ is the length of vectors $\mathbf{m}_i$ and **d**. Following Min et al. (2007) we assume that $\boldsymbol{\Sigma}_m$, the covariance matrix of the model, is equal to the covariance matrix of the observations. Substituting (3) into (2) gives

$$P(\mathbf{m}_i \,|\, \mathbf{d}) = \frac{\exp\left[-\frac{1}{4}(\mathbf{d} - \mathbf{m}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_d^{-1}(\mathbf{d} - \mathbf{m}_i)\right]}{\sum_{j=1}^{N} \exp\left[-\frac{1}{4}(\mathbf{d} - \mathbf{m}_j)^{\mathrm{T}}\boldsymbol{\Sigma}_d^{-1}(\mathbf{d} - \mathbf{m}_j)\right]}. \tag{4}$$

The challenge in calculating (4) arises when trying to invert the covariance matrix $\boldsymbol{\Sigma}_d$. As we must estimate $\boldsymbol{\Sigma}_d$ from observations, the theoretical limit of its condition number is the number of observations (55 yr). This implies that the largest data vector **d**, which we can use to assess model quality, is the length of our observations. If **d** exceeds this, $\boldsymbol{\Sigma}_d$ will be ill-conditioned and we will be unable to calculate the inverse. In practice, a well-conditioned and invertible $\boldsymbol{\Sigma}_d$ requires the length of **d** to be much shorter than the number of observations. This limit on **d** means that, in its current form, (4) cannot be used with high-dimensional geophysical fields as a basis to test model quality. Min et al. (2007) address this by reducing the dimensionality of their model quality metric. We propose a general solution that allows (4) to be calculated from leading eigenvectors of high-dimensional data. Using eigenvector decomposition the inverse of the covariance matrix may be written as

$$\boldsymbol{\Sigma}_d^{-1} = \mathbf{e}\boldsymbol{\Lambda}^{-1}\mathbf{e}^{\mathrm{T}}, \tag{5}$$

where **e** and $\boldsymbol{\Lambda}$ are the eigenvectors and eigenvalues of $\boldsymbol{\Sigma}_d$ (Calenge et al. 2008). Defining $\mathbf{b}_i = \mathbf{e}^{\mathrm{T}}(\mathbf{d} - \mathbf{m}_i)$ and substituting (5) into (3) yields

$$P(\mathbf{d} \,|\, \mathbf{m}_i) = \frac{1}{\sqrt{2\pi^q \det \boldsymbol{\Sigma}_d}} \exp\left(-\frac{1}{4}\sum_k \frac{b_{ik}^2}{\lambda_k}\right), \tag{6}$$

where $\sum_k$ is the summation over the subsequent eigenvectors **e** and eigenvalues $\lambda$ of our observations, ranked in order from largest eigenvalue ($k = 1$) to smallest ($k = K$, where $K$ is the total number of eigenvectors). Substituting (6) into (2) allows us to calculate the model weights as

$$P(\mathbf{m}_i \,|\, \mathbf{d}) = \frac{\exp\left(-\frac{1}{4}\sum_k \frac{b_{ik}^2}{\lambda_k}\right)}{\sum_{j=1}^{N} \exp\left(-\frac{1}{4}\sum_k \frac{b_{ik}^2}{\lambda_k}\right)}. \tag{7}$$

The total number of eigenvectors $K$ is equal to the number of observations. However, in any field with coherent patterns of variability, as $k$ increases the magnitude of the corresponding eigenvalue $\lambda$ decreases. When calculating (7) small eigenvalues cause the exponential term to become very large and the weighting distribution overly selective. Because of this, care must be taken in limiting the number of eigenvalues used to calculate the summation (see below).

### b. Calculating Bayesian weights

We now apply the weighting scheme and assess how the Bayesian-weighting framework can be used to reduce bias in the ensemble average of the CMIP5 historical jet simulations. The data vector **d** is the reanalysis

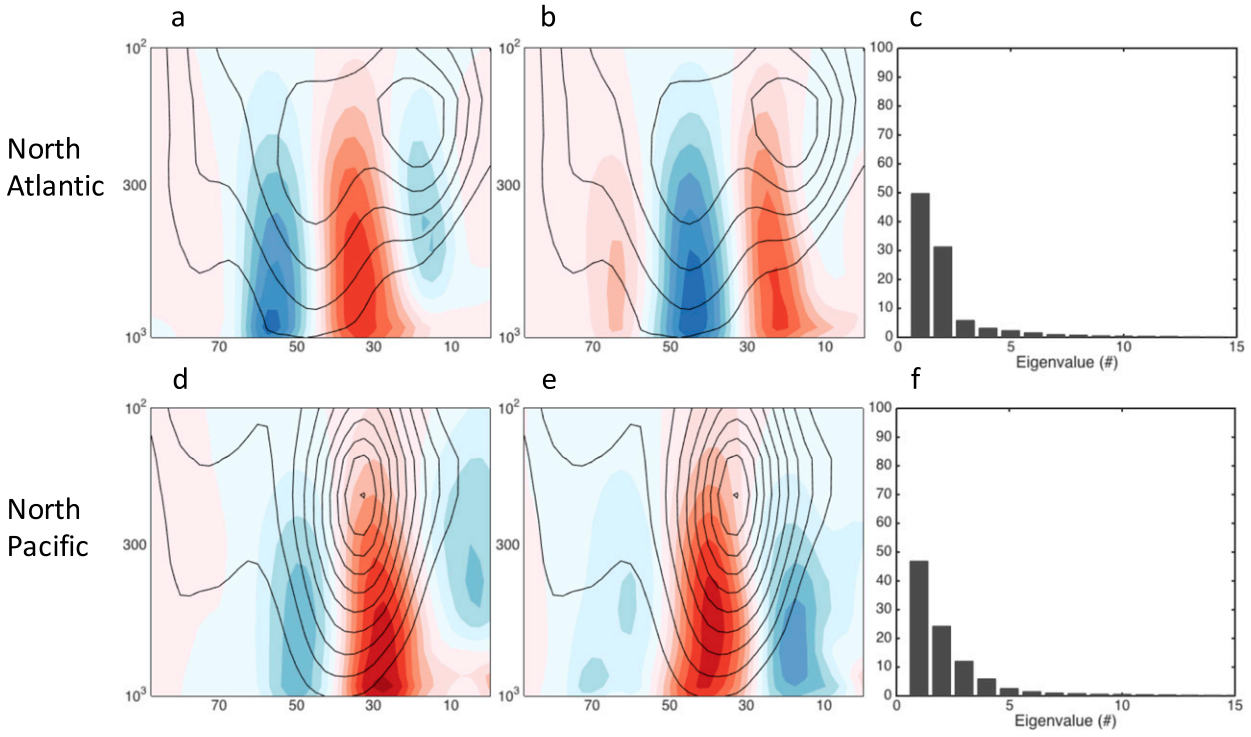


FIG. 4. Leading two eigenvectors of mean DJF density-weighted zonal wind (shading; NCEP–NCAR reanalysis; 1950–2005) over the (a),(b) North Atlantic and (d),(e) North Pacific domains. Black contours show the climatology of NCEP–NCAR zonal wind (see Fig. 2). (c),(f) The percent variance explained by the leading 14 eigenvectors.

climatology shown in Figs. 2a,c but spatially decomposed into an $N \times 1$ vector. The model field we are assessing $\mathbf{m}_i$ is the corresponding $N \times 1$ (spatially decomposed) zonal wind climatology for each simulation. We calculate the eigenvalues and eigenvectors from the $N \times M$ reanalysis data, where $M$ is the number of observations (55 DJF averages between 1950 and 2005).

Figure 4 shows the leading two eigenvectors of density-weighted zonal wind. We apply (7) to the zonal wind field weighted by both density and grid area. Density weighting ensures that the eigenvectors draw power through the depth of the troposphere and not just the upper levels where low-density winds with high variance dominate. Weighting by grid area [cos(lat)] compensates for unequal grid boxes (e.g., Baldwin and Thompson 2009). The first and second eigenvectors in each region represent the north–south wobble and pulsing of the jet, respectively (Lorenz and Hartmann 2003). Together, these patterns explain the majority of variance in the zonal wind field and appear well separated in the eigenvalue spectrum (Figs. 4c,f). The leading six eigenvectors are shown in Figs. S2 and S3 of the supplemental material.

We argue that the dominance of the leading two eigenvectors justifies setting $K = 2$ as the limit in (7). In Fig. 5 we show the distribution of model weights for $K = 2$. This choice is somewhat subjective and depends on inspection of the eigenvalue spectrum and corresponding distributions of weights (shown for $K = 1$–6 in Figs. S4 and S5 of the supplemental material). In the future it would be desirable to formalize this choice using knowledge of sampling errors in EOF analysis (e.g., North et al. 1982; Quadrelli et al. 2005) applied to (7). That ensembles of the same model are assigned consistent weights tells us that the field we are using to assess model quality is characteristic of the model and not simply a result of internal variability. The model weight distribution is different for the North Atlantic and North Pacific, suggesting this metric of model quality is specific to the different phenomena (i.e., a model's ability to simulate the North Atlantic jet is not closely related to its ability to simulate the North Pacific jet). In fact, the Pearson correlation coefficient between the two weight distributions is −0.01. This could be a reflection of the relative importance of the eddy driving and thermal driving in the two jets (Li and Wettstein 2012).

We compare the jets in historical reanalysis to the Bayesian-weighted CMIP5 ensemble average in Fig. 6. While there are still large regions of statistical significance

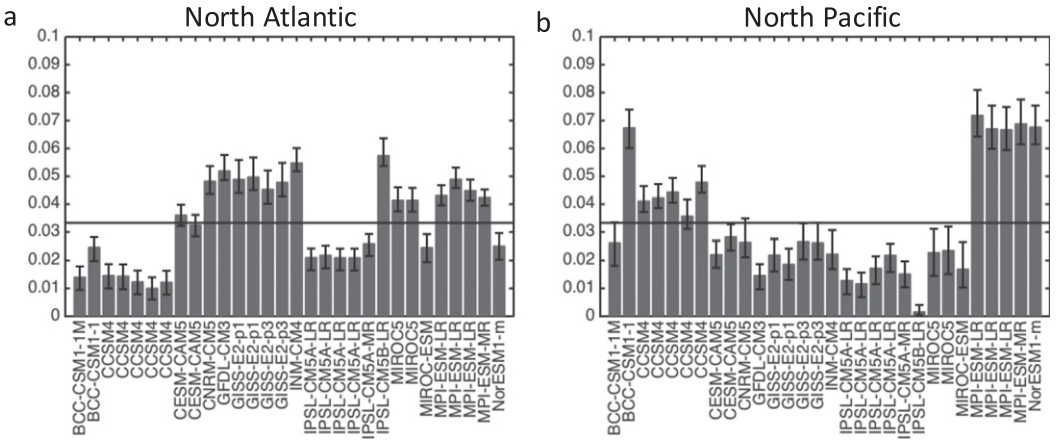


FIG. 5. Distribution of model weights for successive summations of $k$ in (7) for the (a) North Atlantic and (b) North Pacific. Horizontal gray line corresponds to equal model weights $1/N$. Error bars denote a 90% confidence interval calculated by resampling with replacement the years used to estimate the model climatology and reanalysis climatology, eigenvalues, and eigenvectors used in (7). This bootstrapping was repeated 1000 times to produce a distribution of weights for each model.

over the North Atlantic, the magnitude of the Bayesian-weighted (BMA) composite bias is clearly reduced relative to the democratically weighted (AEM) composite shown in Fig. 3. Over the North Pacific the significant equatorward bias is removed completely. This reduced historical bias provides confidence in (7), including our choice of $K$.

## 5. Future jets in CMIP5

We now look at the Bayesian-weighted future climatology of the CMIP5 projections. Figure 7 displays the future jet composites while Fig. 8 shows their difference from historical conditions. Both jets show similar patterns of future change—a strengthening in the upper-tropospheric westerlies between 30° and 50°N and a weakening either side reaching into the mid-to-low troposphere. As mentioned previously, this relative strengthening and narrowing in the zonal mean, time-averaged jet could result from a combination of 1) a physical strengthening and narrowing, 2) reduced variability in latitude position (i.e., less jet "wobbling"), or 3) a reduced latitudinal tilt. Barnes and Polvani (2013) found a shift in the dominant mode of variability in the eddy-driven jet, with less wobbling and more jet pulsing in the future, which is consistent with Fig. 8.

Over the North Atlantic, the upper-level intensification appears to have merged the distinct wind maxima, suggesting this jet may shift toward a combined core regime through midwinter. In idealized modeling studies

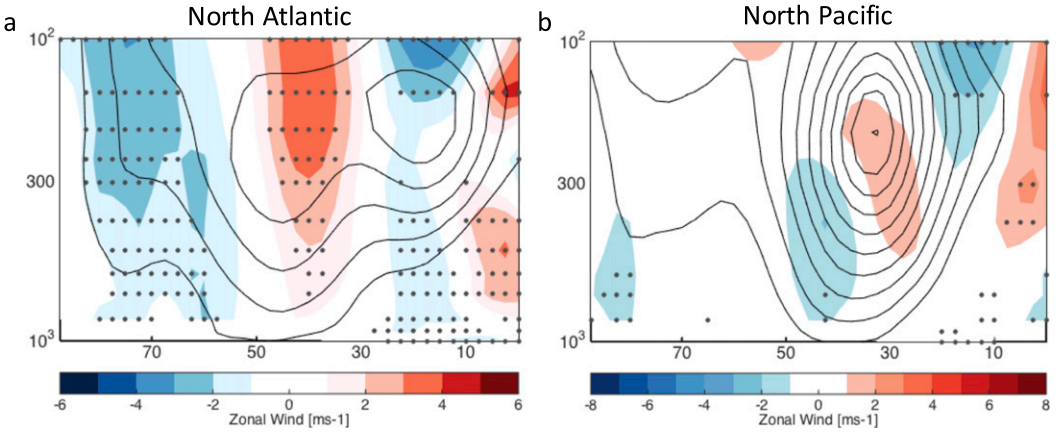


FIG. 6. Difference between NCEP–NCAR reanalysis and CMIP5 BMA zonal wind climatology (1950–2005) for the (a) North Atlantic and (b) North Pacific. Contours show the reanalysis climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$. Stippling marks differences significant at $\alpha_{FDR} = 0.1$ on a two-tailed test (see the appendix).

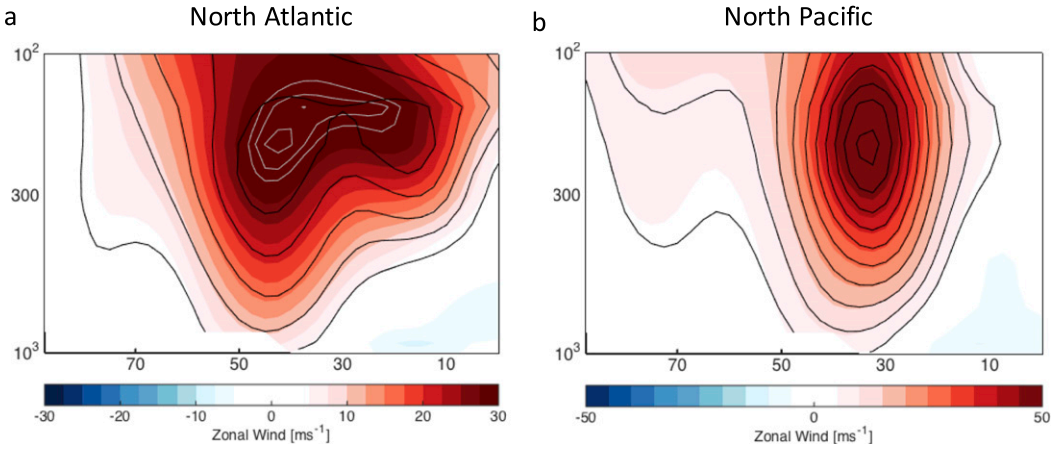


FIG. 7. Future climatology (2075–99) of jets calculated as the Bayesian-weighted ensemble average (shading) for the (a) North Atlantic and (b) North Pacific. White contours show North Atlantic future climatology at intervals of 1 m s$^{-1}$, starting at 30 m s$^{-1}$. Black contours show the historical CMIP5 BMA climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$. Jets are the DJF zonal wind averaged latitudinally across the domains shown in Fig. 1.

Lee and Kim (2003) found that a strong subtropical jet inhibits the latitudinal variability of eddies and the formation of a well-separated (eddy driven) midlatitude jet, which may be what we observe over the North Atlantic in Fig. 7. A poleward jet shift is not the full story of Fig. 8—reduced westerlies poleward of 60°N are coincident with projections of Arctic amplification at these latitudes and appear to be squeezing the poleward flank of the jet. However, it is not possible to conclude whether this is driven by a direct thermal response or is due to changes in eddies.

Consistent with previous studies that use geometric analysis, Fig. 8 shows a poleward shift and intensification at the center of the surface-level eddy-driven jets (e.g., Yin 2005; Barnes and Polvani 2013). The latitude–height cross sections suggest this feature is linked to a significant strengthening of the upper-level westerlies at the jet cores, which is consistent with theoretical expectations of an increased temperature gradient in the upper troposphere (Butler et al. 2010). While our analysis suggests a poleward shift of the eddy-driven jets it is worth considering the statistical robustness of this projection. Although the strengthening and poleward shift toward the surface is associated with a statistically significant response in the upper troposphere, the significance of the northward shift at the surface is limited, particularly in the North Atlantic. This likely represents the competing influence of the equatorward shift associated with low-level Arctic amplification. By virtue of the nonuniform weighting, the standard error constraints (see the appendix) of the Bayesian-weighted changes (Fig. 8) are more stringent than a democratically weighted composite (Fig. S7 in the supplemental material). However, both show this limited significance of the implied surface jet shift.

Figure 9 shows a significant difference between future jet climatology in Bayesian-weighted and democratically weighted composites. These patterns closely match the historical bias (Fig. 3), also indicating that the Bayesian-weighted projected jet change does not differ significantly from the democratically weighted projected jet change (Fig. S7). This suggests that changes in these jets is largely independent of the historical jet characteristics. That is, greenhouse gas (GHG)-forced change in future jets is largely first-order linear. This is in contrast with analysis of CMIP5 ensemble (Bracegirdle et al. 2013) that found the magnitude of the jet shift to be well correlated with the bias in the initial position. Woollings and Blackburn (2012) found a similar lack of coherence between future jet changes and bias in CMIP3 and report that diagnoses of jet stream changes are sensitive to differences in methodology.

## 6. Conclusions

In this study we assess the significant biases in both the North Pacific and North Atlantic jets (Fig. 3). We present a novel Bayesian-weighting framework that reduces this bias (Fig. 6) and use it to assess future changes in the jet structure (Fig. 8). Both regions show broadly similar changes throughout the troposphere; the apparent squeezing of the jet as warming in the upper troposphere pushes the subtropical jet north and Arctic amplification slows westerlies on the jet's poleward flank. That the widely documented poleward shift of the near-surface eddy-driven jet occurs at the confluence of Arctic and tropical influences may explain its lack of statistical robustness during winter. While the upper-tropospheric tropical warming is a well-understood

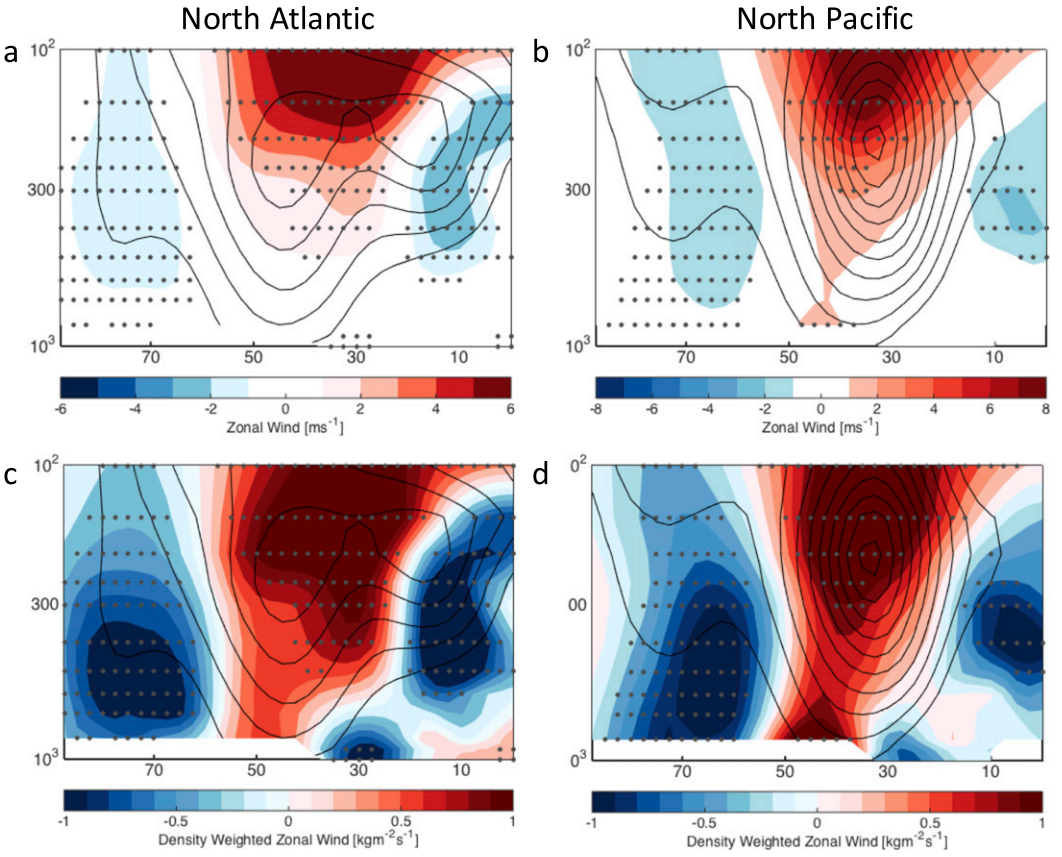


FIG. 8. Projected change in jets calculated as future (2075–99) minus historical (1950–2005) climatology for the (a) and (b). The projected change in jets weighted by density for (c) North Atlantic and (d) North Pacific. Contours show the historical CMIP5 BMA climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$. Stippling marks differences significant at $\alpha_{FDR} = 0.1$ on a two-tailed test (see methods).

response to climate change, the processes involved with Arctic amplification are less clear (Vallis et al. 2015). Quantifying the relative influence of processes such as future sea ice loss (e.g., Screen and Simmonds 2010) and cloud cover (e.g., Li et al. 2015) should allow for more robust predictions of future jet variability. Alternatively, it may be that the processes competing to change jet variability do in fact cancel one another, resulting in no change to surface jet climatology. Over the North Atlantic, the simulations project a shift from a separated to a combined jet regime. This may be a reflection of enhanced control exerted over eddy propagation by the strengthened subtropical jet (Lee and Kim 2003).

Our Bayesian-weighted future jet composites are significantly different from the commonly used democratic mean (AEM; Fig. 9). As such, leveraging the Bayesian composites has the potential to prevent the misinterpretation of future jet conditions. That the Bayesian-weighted future minus historical composites (Fig. 8) are not significantly different from the democratically weighted ones (Fig. S7) suggests that response of the wintertime tropospheric jets to GHG forcing is first-order linear. A significant decision in applying the Bayesian framework was choosing the number of eigenvectors to include in (6). Well-established insight into the physical meaning of the eigenvectors, representing jet wobbling and jet pulsing, and a well-separated eigenvalue spectrum provided a clear choice for $K$. This may not prove to be the case in other applications.

A key goal of this study was to showcase the Bayesian-weighting methodology as applied to high-dimensional fields, which we believe could be useful in a range of applications. While we apply the weighting to the same field used to define model quality (i.e., zonal wind climatology), there is no reason the weights could not be applied to other circulation metrics related to the tropospheric jets. For example, it would be interesting to see whether our weights reduced biases in wave extent, wave speed, and blocking metrics (Barnes and Polvani 2015). A number of recent studies have highlighted the importance of zonal jet asymmetries (Zappa et al. 2013, among others) that are not considered in our analysis of

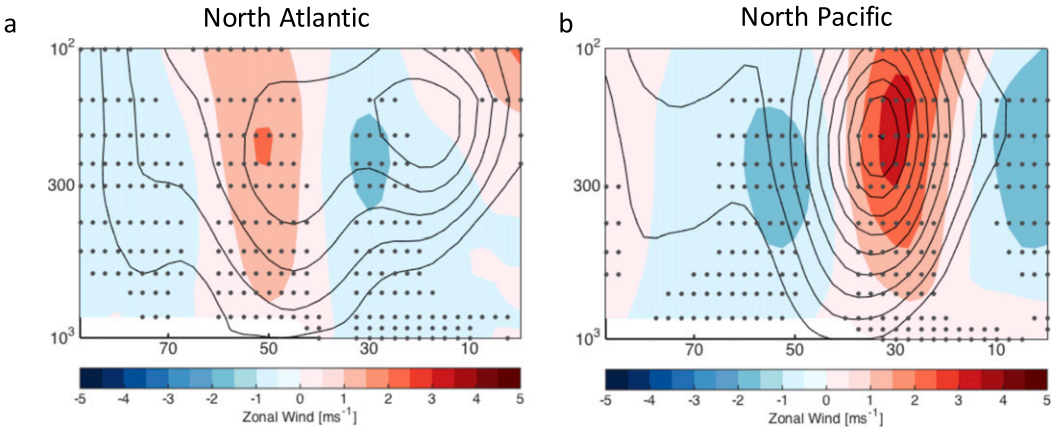


FIG. 9. Difference between projected future (2075–99) climatology calculated using BMA vs AEM (BMA minus AEM) for the (a) North Atlantic and (b) North Pacific. Contours show the reanalysis climatology at intervals of 5 m s$^{-1}$, starting from 5 m s$^{-1}$. Stippling marks differences significant at $\alpha_{FDR} = 0.1$ on a two-tailed test (see methods).

zonal mean climatology. Our primary motivation for studying 2D fields was to illustrate our methodology, but we foresee no reason the approach could not be extended to 3D representations of the zonal wind field. Such analysis would allow the weighting framework to incorporate the bias of overly zonal flow in the quality metric. By choosing to study DJF averages our choice of model quality metric largely disregards some subtleties relating to jet variance. For example, we do not directly consider how faithfully models reproduce the preferred jet locations over the North Atlantic (Woollings et al. 2010) or how this may change in the future.

*Acknowledgments.* DW and DE were supported by the National Science Foundation Grant 1503966. KM was supported by the Ziff Environmental Fellowship from Harvard University's Center for the Environment.

## APPENDIX

### Significance Calculations

All significance levels in this study are calculated from the standard error of the climatologies (mean):

$$\mathrm{SE} = \frac{\Delta}{\sigma}, \tag{A1}$$

where $\Delta$ is the difference between the model and reanalysis climatology, and $\sigma$ is the standard deviation of the climatology. Here $\sigma$ can be calculated as the square root of the variance (see below). In Figs. 3, 6, and 8 the variance of the climatology can be calculated as follows:

$$\mathrm{Var}[E(X)] = \mathrm{Var}\left(\sum_{i=1}^{N} X_i P_i\right) = \mathrm{Var}(X)\sum_{i=1}^{N} P_i^2, \tag{A2}$$

where $E(X)$ is the expected value of zonal wind, which corresponds to climatology. Variable $P_i$ is the probability (weight) of model $i$. In the arithmetic ensemble means this was calculated as $1/N$. Similarly, the variance of the composite means in Fig. 9 was calculated as follows:

$$\begin{aligned}\mathrm{Var}[E(X)_{\mathrm{BMA}} - E(X)_{\mathrm{AEM}}] &= \mathrm{Var}\left(\sum_{i=1}^{N} X_i P_{\mathrm{BMA}i} - \sum_{i=1}^{N} X_i P_{\mathrm{AEM}i}\right) = \mathrm{Var}\left[\sum_{i=1}^{N} (P_{\mathrm{BMA}i} - P_{\mathrm{AEM}i}) X_i\right] \\ &= \mathrm{Var}(X)\sum_{i=1}^{N} (P_{\mathrm{BMA}i} - P_{\mathrm{AEM}i})^2. \end{aligned} \tag{A3}$$

Once the standard error has been used to calculate $p$ values at each grid point, the false discovery rate (FDR) approach is used to calculate critical $p$ values (Ventura et al. 2004; Wilks 2016).

## REFERENCES


Baldwin, M. P., and D. Thompson, 2009: A critical comparison of stratosphere–troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, **135**, 1661–1672, https://doi.org/10.1002/qj.479.

Barnes, E. A., and L. M. Polvani, 2013: Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models. *J. Climate*, **26**, 7117–7135, https://doi.org/10.1175/JCLI-D-12-00536.1.

——, and ——, 2015: CMIP5 projections of Arctic amplification, of the North American/North Atlantic circulation, and of their relationship. *J. Climate*, **28**, 5254–5271, https://doi.org/10.1175/JCLI-D-14-00589.1.

——, and J. Screen, 2015: The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *Wiley Interdiscip. Rev.: Climate Change*, **6**, 277–286, https://doi.org/10.1002/wcc.337.

Bracegirdle, T. J., E. Shuckburgh, J.-B. Sallee, Z. Wang, A. J. Meijers, N. Bruneau, T. Phillips, and L. J. Wilcox, 2013: Assessment of surface winds over the Atlantic, Indian, and Pacific Ocean sectors of the Southern Ocean in CMIP5 models: Historical bias, forcing response, and state dependence. *J. Geophys. Res. Atmos.*, **118**, 547–562, https://doi.org/10.1002/jgrd.50153.

Butler, A. H., D. W. J. Thompson, and R. Heikes, 2010: The steady-state atmospheric circulation response to climate change–like thermal forcings in a simple general circulation model. *J. Climate*, **23**, 3474–3496, https://doi.org/10.1175/2010JCLI3228.1.

Calenge, C., G. Darmon, M. Basille, A. Loison, and J.-M. Jullien, 2008: The factorial decomposition of the Mahalanobis distances in habitat selection studies. *Ecology*, **89**, 555–566, https://doi.org/10.1890/06-1750.1.

Deser, C., G. Magnusdottir, R. Saravanan, and A. Phillips, 2004: The effects of North Atlantic SST and sea ice anomalies on the winter circulation in CCM3. Part II: Direct and indirect components of the response. *J. Climate*, **17**, 877–889, https://doi.org/10.1175/1520-0442(2004)017<0877:TEONAS>2.0.CO;2.

Eichelberger, S. J., and D. L. Hartmann, 2007: Zonal jet structure and the leading mode of variability. *J. Climate*, **20**, 5149–5163, https://doi.org/10.1175/JCLI4279.1.

Francis, J. A., and S. J. Vavrus, 2012: Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophys. Res. Lett.*, **39**, L06801, https://doi.org/10.1029/2012GL051000.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis*. 2nd ed. Chapman and Hall/CRC, 690 pp.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, https://doi.org/10.1029/2007JD008972.

Gong, H., L. Wang, W. Chen, X. Chen, and D. Nath, 2016: Biases of the wintertime Arctic Oscillation in CMIP5 models. *Environ. Res. Lett.*, **12**, 014001, https://doi.org/10.1088/1748-9326/12/1/014001.

Hartmann, D. L., 2007: The atmospheric general circulation and its variability. *J. Meteor. Soc. Japan*, **85B**, 123–143, https://doi.org/10.2151/jmsj.85B.123.

Harvey, B., L. Shaffrey, and T. Woollings, 2014: Equator-to-pole temperature differences and the extra-tropical storm track responses of the CMIP5 climate models. *Climate Dyn.*, **43**, 1171–1182, https://doi.org/10.1007/s00382-013-1883-9.

Hoskins, B. J., and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661–1671, https://doi.org/10.1175/1520-0469(1993)050<1661:RWPOAR>2.0.CO;2.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395–404, https://doi.org/10.1007/s10584-010-9800-2.

——, J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring, 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909–1918, https://doi.org/10.1002/2016GL072012.

Lee, S., and H. Kim, 2003: The dynamical relationship between subtropical and eddy-driven jets. *J. Atmos. Sci.*, **60**, 1490–1503, https://doi.org/10.1175/1520-0469(2003)060<1490:TDRBSA>2.0.CO;2.

Li, C., and J. J. Wettstein, 2012: Thermally driven and eddy-driven jet variability in reanalysis. *J. Climate*, **25**, 1587–1596, https://doi.org/10.1175/JCLI-D-11-00145.1.

Li, Y., D. W. Thompson, and S. Bony, 2015: The influence of atmospheric cloud radiative effects on the large-scale atmospheric circulation. *J. Climate*, **28**, 7263–7278, https://doi.org/10.1175/JCLI-D-14-00825.1.

Liu, J., J. A. Curry, H. Wang, M. Song, and R. Horton, 2012: Impact of declining Arctic sea ice on winter snow. *Proc. Natl. Acad. Sci. USA*, **109**, 4074–4079, https://doi.org/10.1073/pnas.1114910109.

Lorenz, D. J., and D. L. Hartmann, 2003: Eddy–zonal flow feedback in the Northern Hemisphere winter. *J. Climate*, **16**, 1212–1227, https://doi.org/10.1175/1520-0442(2003)16<1212:EFFITN>2.0.CO;2.

Min, S. K., A. Hense, H. Paeth, and W. T. Kwon, 2004: A Bayesian decision method for climate change signal analysis. *Meteor. Z.*, **13**, 421–436, https://doi.org/10.1127/0941-2948/2004/0013-0421.

——, D. Simonis, and A. Hense, 2007: Probabilistic climate change predictions applying Bayesian model averaging. *Philos. Trans. Roy. Soc. London*, **365A**, 2103–2116, https://doi.org/10.1098/rsta.2007.2070.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

Quadrelli, R., C. S. Bretherton, and J. M. Wallace, 2005: On sampling errors in empirical orthogonal functions. *J. Climate*, **18**, 3704–3710, https://doi.org/10.1175/JCLI3500.1.

Screen, J. A., and I. Simmonds, 2010: The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, **464**, 1334–1337, https://doi.org/10.1038/nature09051.

Shaw, T. A., and Coauthors, 2016: Storm track processes and the opposing influences of climate change. *Nat. Geosci.*, **9**, 656–664, https://doi.org/10.1038/ngeo2783.

Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.*, **7**, 703–708, https://doi.org/10.1038/ngeo2253.

Simpson, I. R., T. A. Shaw, and R. Seager, 2014: A diagnosis of the seasonally and longitudinally varying midlatitude circulation response to global warming. *J. Atmos. Sci.*, **71**, 2489–2515, https://doi.org/10.1175/JAS-D-13-0325.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Vallis, G. K., P. Zurita-Gotor, C. Cairns, and J. Kidston, 2015: Response of the large-scale structure of the atmosphere to

global warming. *Quart. J. Roy. Meteor. Soc.*, **141**, 1479–1501, https://doi.org/10.1002/qj.2456.

van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: An overview. *Climatic Change*, **109**, 5, https://doi.org/10.1007/s10584-011-0148-z.

Ventura, V., C. J. Paciorek, and J. S. Risbey, 2004: Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. *J. Climate*, **17**, 4343–4356, https://doi.org/10.1175/3199.1.

Wilks, D. S., 2016: ''The stippling shows statistically significant grid points'': How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

Wittman, M. A. H., A. J. Charlton, and L. M. Polvani, 2005: Notes and correspondence on the meridional structure of annular modes. *J. Climate*, **18**, 2119–2122, https://doi.org/10.1175/JCLI3394.1.

Woollings, T., and M. Blackburn, 2012: The North Atlantic jet stream under climate change and its relation to the NAO and EA patterns. *J. Climate*, **25**, 886–902, https://doi.org/10.1175/JCLI-D-11-00087.1.

——, A. Hannachi, and B. Hoskins, 2010: Variability of the North Atlantic eddy-driven jet stream. *Quart. J. Roy. Meteor. Soc.*, **136**, 856–868, https://doi.org/10.1002/qj.625.

Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.*, **32**, L18701, https://doi.org/10.1029/2005GL023684.

Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, **26**, 5379–5396, https://doi.org/10.1175/JCLI-D-12-00501.1.