Disentangling forced trends in the North Atlantic jet from natural variability using deep learning

Alejandro Hermoso[1], **Sebastian Schemm**[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland

Key Points:

- A deep learning-based method is able to extract forced trends in the winter North Atlantic jet stream from climate simulations.
- The characteristics of the forced trends are consistent across different periods.
- Forced trends show an intensification of the jet along a southwest-northeast oriented band and a jet extension over Europe.

Corresponding author: Alejandro Hermoso, alejandro.hermoso@env.ethz.ch

Abstract

Regional weather variability and extremes over Europe are strongly linked to variations in the North Atlantic jet stream, especially during the winter season. Projections of the evolution of the North Atlantic jet are essential for estimating the regional impacts of climate change. Therefore, separating forced trends in the North Atlantic jet from its natural variability is an extremely relevant task. Here, a deep learning based method, the Latent Linear Adjustment Autoencoder (LLAE), is used for this purpose on an ensemble of fully-coupled climate simulations. The LLAE is based on an autoencoder and an additional linear component. The model predicts the wind component affected by natural variability by using detrended temperature and geopotential as inputs. The residual between this prediction and the original wind field is interpreted as the forced component of the jet. The method is first tested for the geostrophic wind for which the forced trend can be obtained analytically from the difference between geostrophic wind computed from detrended and full geopotential. Despite the large variability of the original trends, the LLAE is shown to be effective in extracting the forced component of the trend for each individual ensemble member in both geostrophic and full wind fields. The LLAE-derived forced trend shows an increase in the upper-level zonal wind speed along a southwest-northeast oriented band over the ocean and a jet extension towards Europe. These are common characteristics over different periods and show some similarities to the upper-level zonal wind speed trend obtained from the ERA5 reanalysis.

Plain Language Summary

The North Atlantic jet stream, which is a narrow and strong air current flowing from west to east, has a substantial influence on daily weather variability and temperature and precipitation extremes in Europe. It is therefore essential to investigate how the North Atlantic jet is being modified under climate change. Trends in the North Atlantic jet are the combination of an external forcing produced by the increase in greenhouse gases and natural variability across multiple time scales. In this study, we use a data driven model called Latent Linear Adjustment Autoencoder (LLAE) to separate the jet trends related to the effect of the increase of greenhouse gases from the natural variability. The output of a set of climate simulations, namely zonal wind, temperature and geopotential, is used to adapt the adjustable parameters of the LLAE. After this step, the difference between the wind provided by the LLAE and the original field contains the effect of the external forcing. The method is successful in removing most effects associated with natural variability and reveals an intensification increase of the North Atlantic jet along a southwest-northeast oriented band and an extension over Europe.

1 Introduction

Jet streams play an important role in driving regional weather variability and generating extreme events. In particular, variations in the position and intensity of the North Atlantic jet stream have a strong impact on temperature and precipitation over Europe, especially during winter (Hurrell, 1995; Deser et al., 2017). Climate simulations provide an estimate of the changes in the position and intensity of the jet stream produced by an increase of greenhouse gases. However, there is a large uncertainty in these future projections (Simpson et al., 2014; McKenna & Maycock, 2021) and generally they fail to capture the observed strengthening of the North Atlantic jet (Blackport & Fyfe, 2022). The North Atlantic jet stream exhibits substantial variability on short time scales due to transitions among different regimes (Novak et al., 2015), as well as on longer time scales, influenced by tropical variability on sub-seasonal scales produced by the Madden-Julian Oscillation (Yadav & Straus, 2017) or on a decadal scale affected by El Niño-Southern Oscillation (Schemm et al., 2018).

In this context, disentangling forced trends in the North Atlantic jet, induced by, for example, higher concentrations of greenhouse gases, from natural variability constitutes a challenging task. The forced response can in principle be identified by averaging across a large ensemble of climate simulations with slightly different initial conditions (Deser et al., 2012; Bengtsson & Hodges, 2019; Dai & Bloecker, 2019). The large-ensemble approach is computationally demanding and is unfeasible when the objective is to identify forced trends in observational data as in this case only one realization of the system (i.e., the observed time series) is available. In this regard, dynamical adjustment techniques aim at separating forced trends from natural variability (Guan et al., 2015; Lehner et al., 2017; Terray, 2021; Deser & Phillips, 2023). These methods are even applicable to small ensembles or observational records. Previous works based on these techniques use a circulation proxy to estimate the natural variability of a variable of interest, such as temperature or precipitation. Linear regression (Smoliak et al., 2015; de Vries et al., 2023) or analog methods (Deser et al., 2016) are further popular strategies, although their application is limited to large temporal and spatial scales where nonlinearities are less important.

A new approach to disentangle forced trends from natural variability is the Latent Linear Adjustment Autoencoder (LLAE), which is based on deep learning (Heinze-Deml et al., 2021). The method combines a variational autoencoder (Kingma & Welling, 2021), consisting of an encoder and a decoder, whose architecture consists of multiple convolutional neural networks (LeCun et al., 2015), with a linear model. The encoder translates the field of interest (e.g., precipitation) into a low-dimensional latent space, which is subsequently decoded for reconstruction of the original field. During the training process, the decoder is penalized if it does not adequately reconstruct the original field (e.g., precipitation) from the latent space, based on the reconstruction loss. The linear model uses a proxy variable (e.g., detrended sea-level pressure), assumed to be independent from the external forcing – through detrending – as input and is trained to predict the latent space.

In the original work of Heinze-Deml et al. (2021) both model components of the latent linear autoencoder are trained sequentially. Afterwards, the encoder part is removed and the latent space is predicted from the proxy variable with the external forcing removed (e.g., detrended sea-level pressure). The decoder reconstructs the associated precipitation from the latent space and the result is the circulation-induced precipitation free of the external forcing. The precipitation residual, which is the difference between the reconstructed precipitation from detrended sea level pressure (SLP) and the simulated precipitation, is the part of the precipitation attributed to the external forcing. In other words, the LLAE is thus interpreted as a statistic model that reconstructs the precipitation field from detrended sea-level pressure. The advantages of this method over other dynamical adjustment strategies are the non-linearity introduced by the decoder and the possibility to directly obtain the forced response over the whole area of interest instead of gridpoint-wise. Furthermore, as shown in Heinze-Deml et al. (2021), this technique can effectively detect forced trends at the regional scale even when the ensemble size is considerably reduced, providing similar results to the mean of a larger ensemble. However, a key problem remains related to whether the removal of the linear trend is sufficient to assume that the input of the LLAE is independent of the external forcing.

Previous studies used dynamical adjustment techniques to detect forced trends in precipitation (Smoliak et al., 2015; Lehner et al., 2017; Guo et al., 2019; Sippel et al., 2019; Heinze-Deml et al., 2021). However, the main objective of this work is to identify the forced trend in the North Atlantic winter jet from a small ensemble of fully-coupled climate simulations. A related study investigated this problem using a deep learning approach applied to idealized climate simulations performed with a dry dynamical core (Connolly et al., 2023). In our work, the data are provided by a small ensemble of fully-coupled cli-

mate simulations including all processes. Daily upper-level zonal wind speed is the variable of interest, and detrended upper-level temperature and geopotential height are used as input to the linear component of the model as the trend-free variables from which the decoder reconstructs a wind field.

Apart from the scientific question of how to separate the forced trend from the natural variability, the wind speed is ideal for testing the LLAE strategy, as the geostrophic wind can be directly computed based on (detrended) geopotential. In this case, the difference between the geostrophic wind computed with the full geopotential and the geostrophic wind obtained from the detrended geopotential is interpreted as the component of the geostrophic wind that is related to the trend in geopotential. The LLAE results can then be compared to the analytical results. This is not possible for the relationship between SLP and precipitation used in Heinze-Deml et al. (2021).

The structure of the paper is as follows. Section 2 describes the LLAE and the climate simulations. Section 3 discusses the forced trends produced by the LLAE for both geostrophic and full wind for different periods. Main conclusions are provided in section 4.

2 Data and Methods

2.1 Latent Linear Adjustment Autoencoder

The statistical method applied to identify forced trends in the jet stream follows the procedure of Sippel et al. (2019) and Heinze-Deml et al. (2021). Given a variable of interest Y, affected by both natural variability and external forcing, and an input X, which is assumed to be unaffected by the external forcing, the component of Y represented by X, $\hat{Y}_X$ can be generally described by:

$$\hat{Y}_X = f(X), \tag{1}$$

where f is a nonlinear function. In this case, Y is a two-dimensional zonal wind field and X is a vector of input features containing detrended geopotential height and temperature. Further details on the climate simulation data are provided in section 2.3. If the input X is unaffected by external forcing, the residual $\hat{R} = Y - \hat{Y}_X$ between the component explained by the input, $\hat{Y}_X$ and the original field Y contains the effect of the forcing on Y. However, any signal that is not explained by the input X is part of the residual, which could mask the interpretation of the impact of external forcing if the choice of the input variables X is inappropriate, so that a large part of the output is unexplained by the input. The incorporation of geopotential height in the input features in combination with temperature in our application is intended to minimize this issue. In order to remove the effects of the external forcing on the input X, these variables are linearly detrended.

The function f is approximated trough the LLAE developed in Heinze-Deml et al. (2021). Figure 1 provides a schematic of the full method. The model combines an autoencoder with a linear model to estimate $\hat{Y}_X$. The autoencoder (upper row in Fig. 1a) consists of an encoder (e) that maps the original field Y into a low-dimensional latent space. Then, the decoder (d) maps the latent space to a reconstruction of the original field, $\hat{Y}$. Following the original work, the dimension of the latent space is set to 400. The architecture of both the encoder and the decoder consists of three convolutional neural networks with a kernel size of 3 and filter sizes of 16, 32 and 64, and a residual layer. In addition to the autoencoder, a linear component (l) is included to predict the latent space from the input X (lower row in Fig. 1). The autoencoder and the linear component are alternately trained. First, the autoencoder parameters are updated, keeping the parameters of the linear component fixed. The loss function contains two terms, one for the

autoencoder and another one for the linear model.

$$L_1 = L_{AE} + \lambda L_l \tag{2}$$

Following Heinze-Deml et al. (2021), the first term, L_{AE}, measures the difference between the reconstruction and the input and the Kullback–Leibler divergence between the distribution of the encoded inputs and the prior distribution of the latent space, which is chosen to be a standard multivariate Gaussian distribution. The second term, L_l, includes the difference between the original field, Y and the decoded outputs based on the latent space obtained from X, Y_X:

$$L_l = ||Y - d(l(X))||^2 \tag{3}$$

For the experiments presented here, $\lambda = 1$, so that both terms contribute equally to the loss function.

In the second step, the parameters of the linear model are adjusted, keeping the autoencoder parameters fixed, by using a loss function that measures the difference between the latent spaces obtained from the encoder and from the linear model.

$$L_2 = ||e(Y) - l(X)||^2 \tag{4}$$

This procedure links the parameters of the autoencoder, which provides the prediction $\hat{Y}_X$, to those of the linear model, which computes the latent space used by the decoder from the input X. The model is trained with the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 10^{-3} for 65 epochs. After training, the encoder is no longer needed and only the input data X is used to obtain the latent space variables through the linear component. Then, the decoder produces a prediction of the field Y, which is assumed be unaffected by external forcing, $\hat{Y}_X$.

Predictions $\hat{Y}_X$ are evaluated by means of the R^2, which indicates the proportion of explained variance by the predictions, and the mean squared error (MSE) for one test member. In addition, reconstructions ($\hat{Y}$) and predictions are shown for various values of the distribution of R^2 to provide insight on the performance of the LLAE and discuss the errors on poor predictions in terms of explained variance.

The residual $\hat{R}$ is used to estimate forced trends. In particular, a linear trend of $\hat{R}$ is computed for different periods, the full period included in the dataset (1980–2100) and a present-day period (1980–2022). Note that the residual includes all effects on Y (zonal wind) that are not explained by the input X (detrended geopotential height and temperature). Therefore, a fraction of the trend obtained from the residual could be attributed to additional processes, not exclusively to external forcing. Detrended inputs are used for both training and inference. The method is applied to both geostrophic and full wind. If both trends are similar, it can be interpreted that changes in the pressure gradient force are the main factor driving forced trends in the full wind. Otherwise, other terms contributing to the ageostrophic component of the wind would be relevant to produce forced changes. In addition, the application of the LLAE to geostrophic wind allows us to prove the efficiency of the method on a field for which there is a simple alternative to compute forced trends.

2.2 Forced trends in the geostrophic wind

Before using the LLAE to estimate the forced trends in the full zonal wind speed at 250 hPa, we compute the forced trends in the geostrophic wind by analyzing the difference between the geostrophic wind computed from non-detrended and detrended geopotential at the same pressure level. Here, it is assumed that the detrending procedure is effective in removing the forced trend and that the detrended geopotential contains only

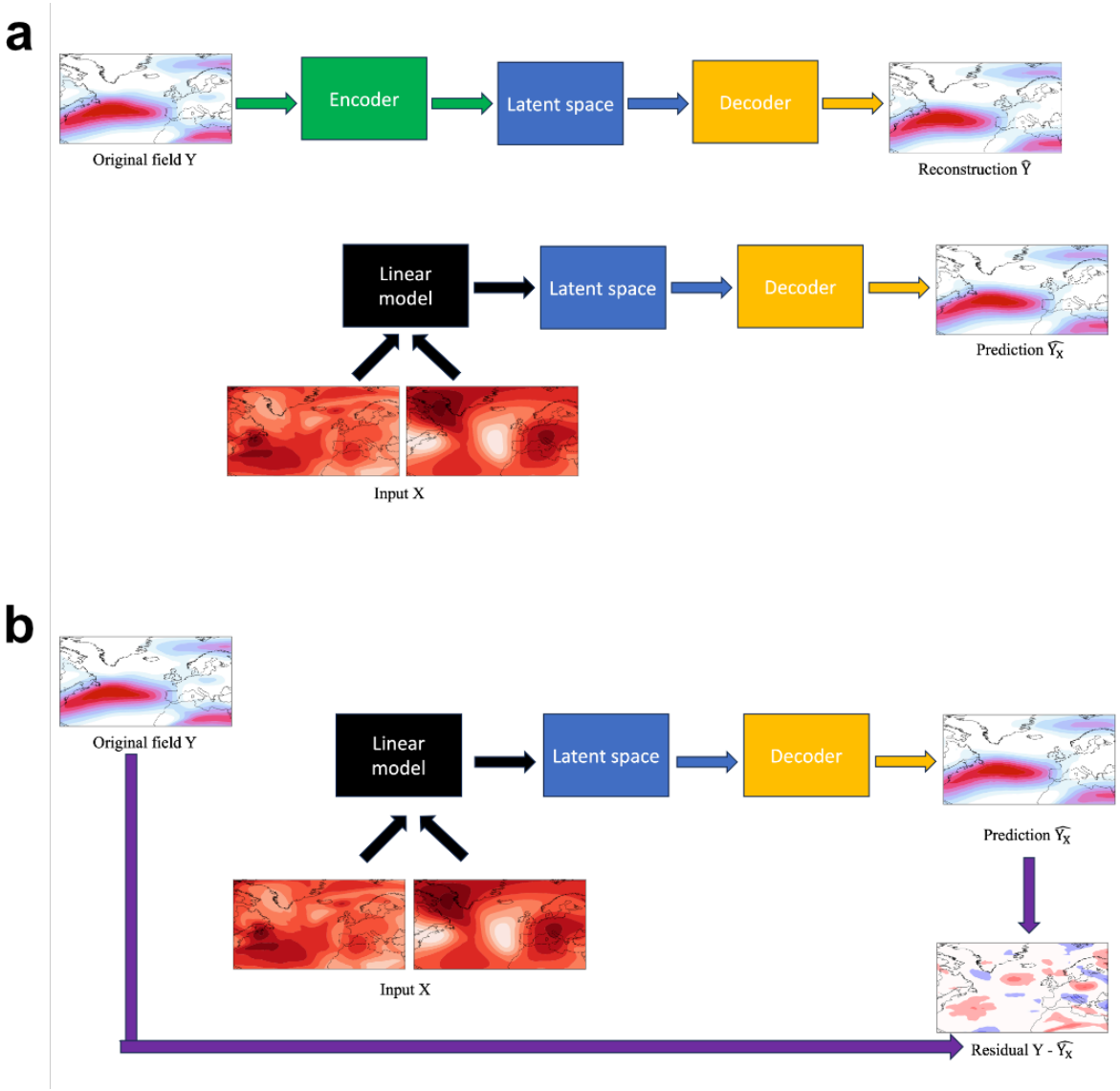

Figure 1: Schematic of the Latent Linear Adjustment Autoencoder. a) The autoencoder component of the model, which consists of multiple convolutional layers for both the encoder and the decoder (see text for details), encodes the field of interest Y into a low-dimensional latent space. The decoder transforms this intermediate output back to the original space to obtain a reconstruction $\hat{Y}$. The linear component uses an input X, which is assumed to be independent of the external forcing, to predict the latent space and obtain the reconstruction $\hat{Y}_X$ from it. During training, the parameters of the autoencoder and those of the linear model are updated in alternative steps with the goal of reducing the difference between $\hat{Y}$ and Y and between the latent spaces computed from the encoder and the linear model. b) Once the training is completed, the encoder is no longer needed and the latent space is predicted from the input X. The residual between the reconstruction obtained from X, $\hat{Y}_X$ and the original field Y is interpreted as the part of Y that is independent of the external forcing.

the natural variability[1]. This gives an indication of changes in the zonal wind due to forced changes in the pressure gradient force and provides a baseline to analyze the performance of the LLAE.

2.3 Climate simulation data

We apply the LLAE to an ensemble of 5 climate simulations performed with the Community Earth System Model (CESM), version 2.1.2, labeled 0900, 1000, 1100, 1200 and 1300. The model is run in fully coupled mode including the Community Atmosphere Model (CAM6) (Bogenschutz et al., 2018; Danabasoglu et al., 2020) with 32 vertical levels), the Community Land Model (CLM5) (Lawrence et al., 2019), the Parallel Ocean Program version 2 (POP2, 60 vertical levels), the Los Alamos National Laboratory Sea Ice model (CICE5) (Hunke et al., 2015), and the hydrological routing model Model for Scale Adaptive River Transport (MOSART) (Li et al., 2013). The period with prescribed forcing covers from 1850 to 2014, and from 2015 to 2100 the runs are forced with the SSP3-7.0 scenario, which constitutes a medium-to-high forcing pathway (O'Neill et al., 2016). The horizontal resolution is approximately 1° and output is available every 6 hours.

The target variables are the geostrophic and full zonal wind speed at 250 hPa over the North Atlantic during the winter season (DJF). The chosen domain extends between 20°N and 80°N and between 80° W and 35°E and consists of 64 × 93 grid points. This choice ensures that the eddy-driven jet stream is well captured, while reducing the influence of subtropical jet. The 6-hourly data are aggregated to daily means. The input data X consists of linearly detrended geopotential height and temperature at the same pressure level, obtained from the aforementioned CESM climate simulations. Therefore, the input vector has a length of 11904 ($64 \times 93 \times 2$). Fig. 2 shows 250 hPa temperature and geopotential height time series averaged over the entire domain and the respective detrended time series. The detrending procedure does not affect variability on shorter time scales, such as decadal variability.

Data between 1980 and 2100 of three ensemble members are used as training data, while the remaining ensemble members are used as a test dataset. Given the relatively small ensemble size and thus the limited amount of available data for training, the model is trained independently on all possible combinations of three ensemble members. The results shown in the next section represent the average of the test members across the different combinations. To avoid stability problems during training, input and target data are normalized. Wind speed values are divided by the 99.9 percentile of the training data and input data are normalized using the 0.1 and 99.9 percentiles. Outputs are rescaled back to the original units to compute trends and error measures, such as the MSE.

[1] We acknowledge that this assumption is highly debated and various detrending methods exist. For the purpose of this study linear detrending seems sufficient

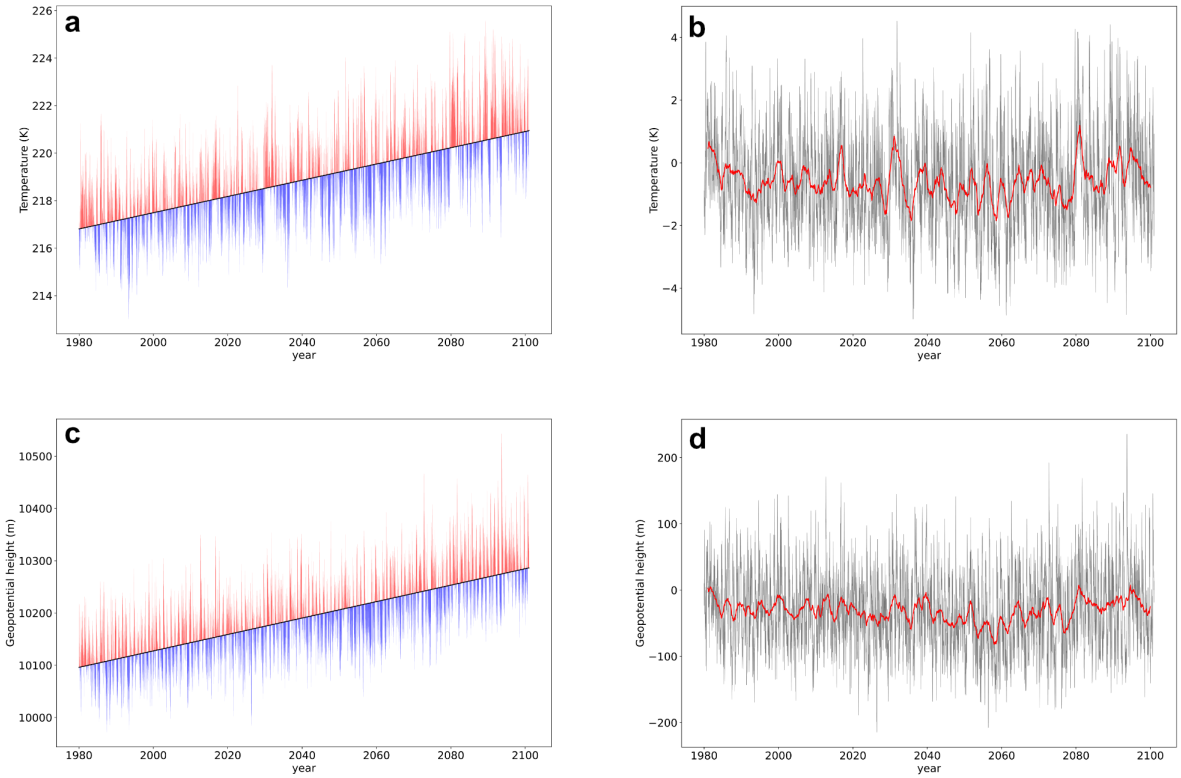

Figure 2: Time series of a) 250-hPa temperature and c) 250-hPa geopotential height for the ensemble member 0900 averaged over the entire domain (20–80°N, 80°W–35°E). The black line represents the linear trend for the whole simulation period (1980-2100). Temperature and geopotential height values are represented in red (blue) if they are above (below) the linear trend. Panels b) and d) represent the time series of detrended 250-hPa temperature and geopotential height centered around zero (gray line) and yearly running mean (red line).

3 Results

3.1 Forced trends in the geostrophic wind: Analytic estimate

In order to have a reference to analyze the forced trends obtained from the LLAE, we first examine the original and analytically computed forced trends in the geostrophic wind. A large spread characterizes the full trend in the zonal geostrophic wind, and therefore the direction of the jet shift is not consistent across the ensemble (Fig. 3, left column). The forced trend in the geostrophic wind, which is the trend in the residual between the geostrophic wind computed from geopotential and detrended geopotential, for all ensemble members in the present-day period (1980–2022)[2] shows an equatorward shift of the zonal wind over the western North Atlantic and a jet extension over Europe (Fig. 3 right column). The estimated forced trend is fairly similar for all ensemble members with a positive trend along a southwest-northeast oriented band, despite the large ensemble spread that characterizes full trends in the geostrophic wind. This indicates that this approach is capable of removing a large part of the natural variability responsible for the high ensemble spread, although some differences between the members remain, especially over the northwestern sector of the domain.

The variability among ensemble members in the full simulation period (1980–2100) is reduced compared with the present day period (1980–2022), in the original trends (Fig. 4, left column). The trend patterns are now broadly consistent with an equatorward shift over the western North Atlantic and a downstream extension over Europe seen in all ensemble members. Forced trends for this period (Fig. 4, right column) are also characterized by an increase in geostrophic wind along a southwest-northeast oriented band. Differences between the ensemble members remain over the Labrador Sea region. While the member-to-member variability was still large during the present period, the estimated forced trend is as before remarkably similar between all members and both periods. The fact that the full trend converges to the forced trend computed over the entire simulation period lends credence to the method of estimating the forced trend based on even the present-day period.

3.2 Evaluation of the LLAE

Before we anaylze the North Atlantic winter jet, consideration is given to the global performance of the model through R^2 and MSE. A large fraction of the variance, above 70 %, is explained across most of the domain, except for an area southeast of Greenland (Fig. 5a). This indicates that the input used is adequate to obtain satisfactory predictions of the zonal wind affected by natural variability. In this case, the residual can be safely interpreted as the forced wind component. The MSE is generally low, below 0.5 $\mathrm{m\,s^{-1}}$, except over some areas in the North Atlantic over the Gulf Stream, the south eastern tip of Greenland and at the end of the storm track, where errors are slightly larger (Fig. 5b). These are regions where day-to-day variability is high, also due to frequent explosive cyclogenesis in these regions. In summary, these results show a general good performance of the LLAE to reconstruct the zonal wind from temperature and geopotential, which suggests a low likelihood of attributing forced signals to spurious effects. Next we look at individual cases of well- and poorly-reconstructed flow situations.

Fig. 6 shows multiple reconstructions, obtained from the application of the autoencoder to the original wind field, and predictions, obtained from the combination of the linear model and the decoder using detrended temperature and geopotential as input, for different values of R^2. Poorly reconstructed (during training) and predicted (from temperature and geopotential) flow fields are situations with cut-off formation over the

[2] Here present-day period refers to years between the start of the simulations (1980) and 2022, although the SSP3-7.0 emission scenario is applied from 2015.

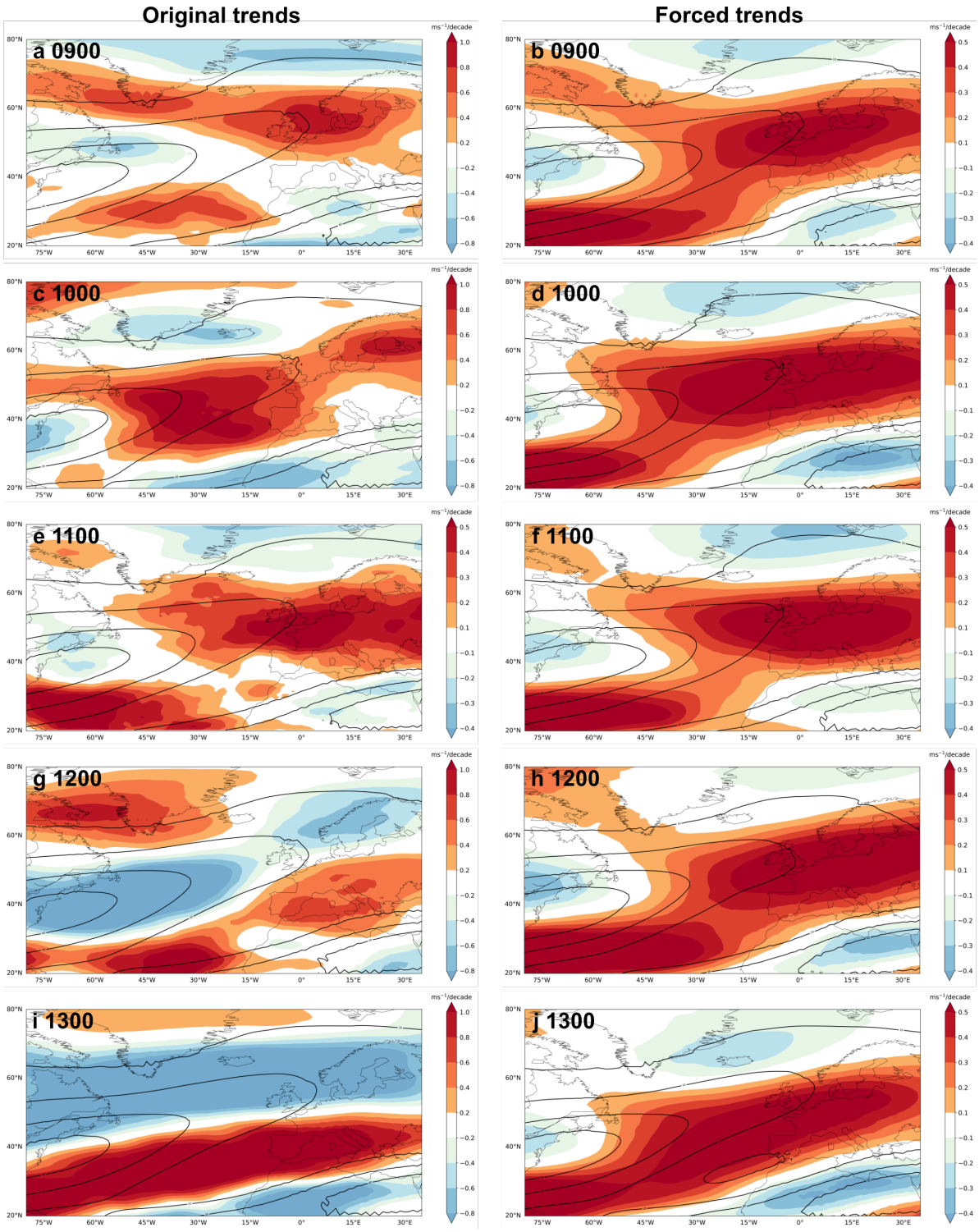

Figure 3: Left column: Linear trends of the geostrophic zonal wind speed at 250 hPa for each ensemble member (shading) and climatological mean (black contours, between 10–40 $\mathrm{m\,s^{-1}}$) for the present-day period 1980–2022. Right column: Linear trend of the difference between geostrophic zonal wind computed from non-detrended and detrended geopotential (shading) and climatological mean of the geostrophic zonal wind (black contours, between 10 and 40 $\mathrm{m\,s^{-1}}$). Note that the range of the color scale for the original trends is the double of the range of the forced trends.

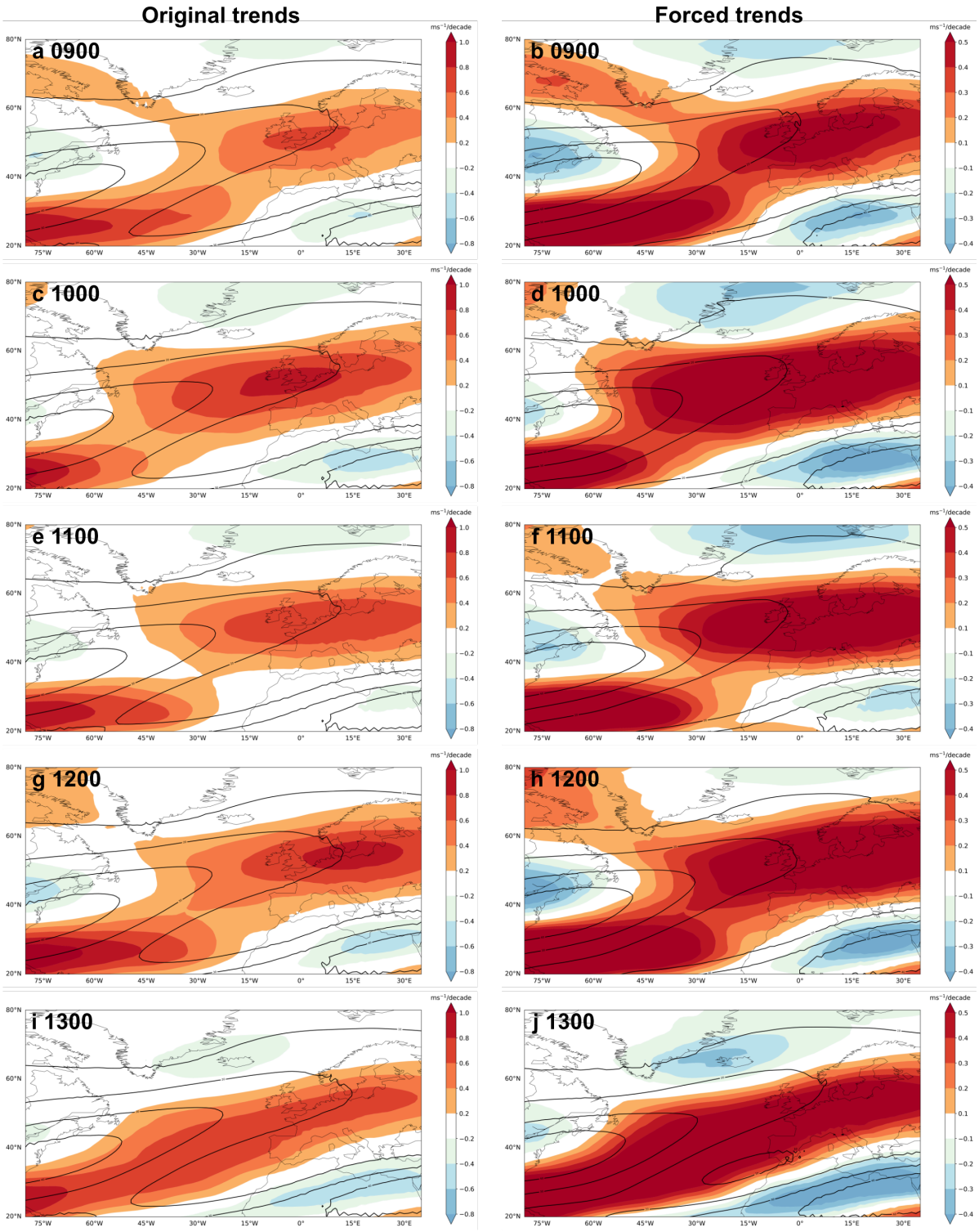

Figure 4: As in Fig. 3, but for the full simulation period 1980–2100.

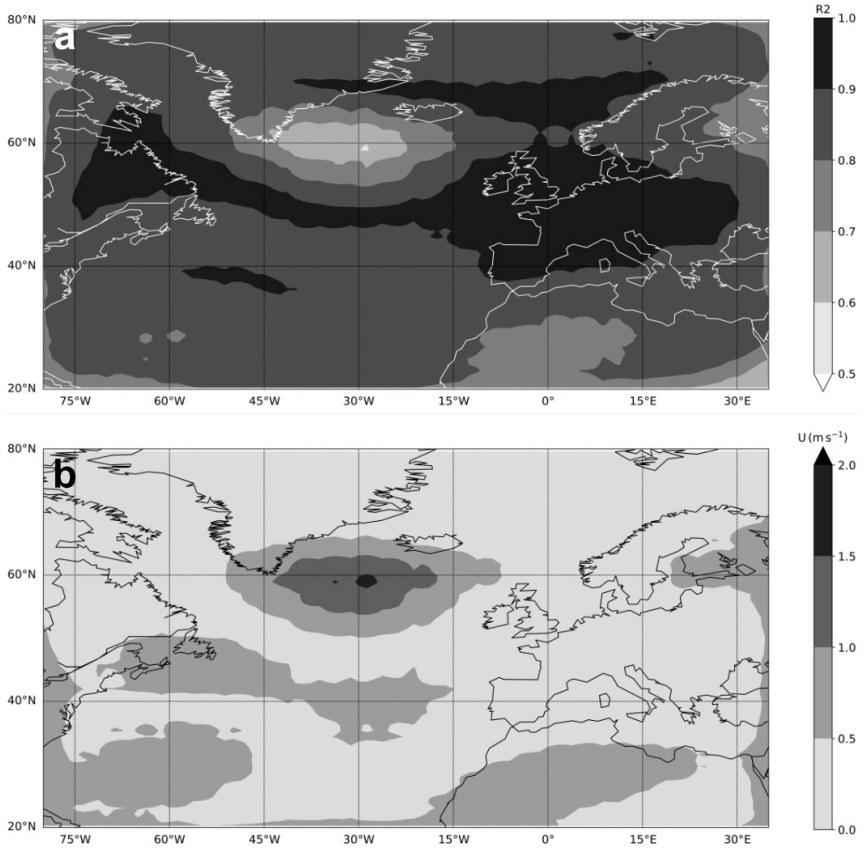

Figure 5: Evaluation metrics for 250-hPa zonal wind predictions obtained from the LLAE: a) R^2 (proportion of explained variance) and b) MSE.

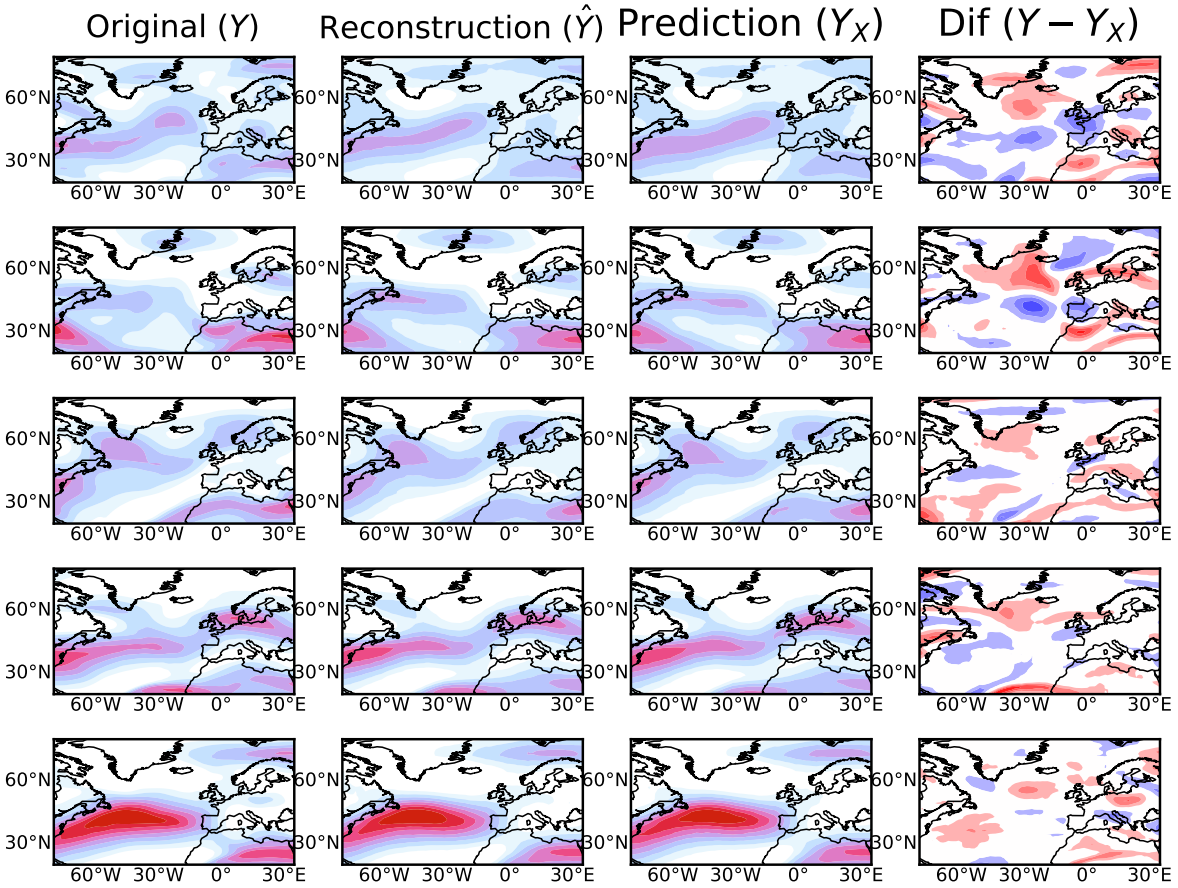

Figure 6: Left to right: Original zonal wind fields at 250 hPa, Y, reconstructions, $\hat{Y}$, predictions, $\hat{Y}_X$, and difference between original fields and predictions, Y - $\hat{Y}_X$ for different values of R^2. Top to bottom: worst example, percentiles 25, 50, 75 of the R^2 distribution and best example.

North Atlantic. Poorer performance is also obtained for weaker jets, while the LLAE achieves better predictions in terms of R^2 for stronger, less wavy jets with a more zonal orientation that dominate over the North Atlantic.

3.3 Quantifying the forced trends

The difference between the predictions obtained from the LLAE and the original geostrophic or full zonal wind speed are used to estimate forced trends. Consideration is given to the the present-day period (1980–2022) and full simulation period (1980–2100).

Forced trend in geostrophic wind We start with the forced trends in the geostrophic wind using the LLAE. The forced trends (after dynamical adjustment) of the five ensemble members are characterized by an increase of the geostrophic wind along a southwest-northeast oriented band, an extension of the jet towards Europe and an equatorward shift over the western North Atlantic, which generates a reduction of the wind speed near the Gulf Stream region (Fig. 7). Forced trends show some variability at low latitudes over the eastern North Atlantic and north Africa and a considerable spread at high latitudes, especially over southern Greenland and the Labrador sea. Over these areas, some par-

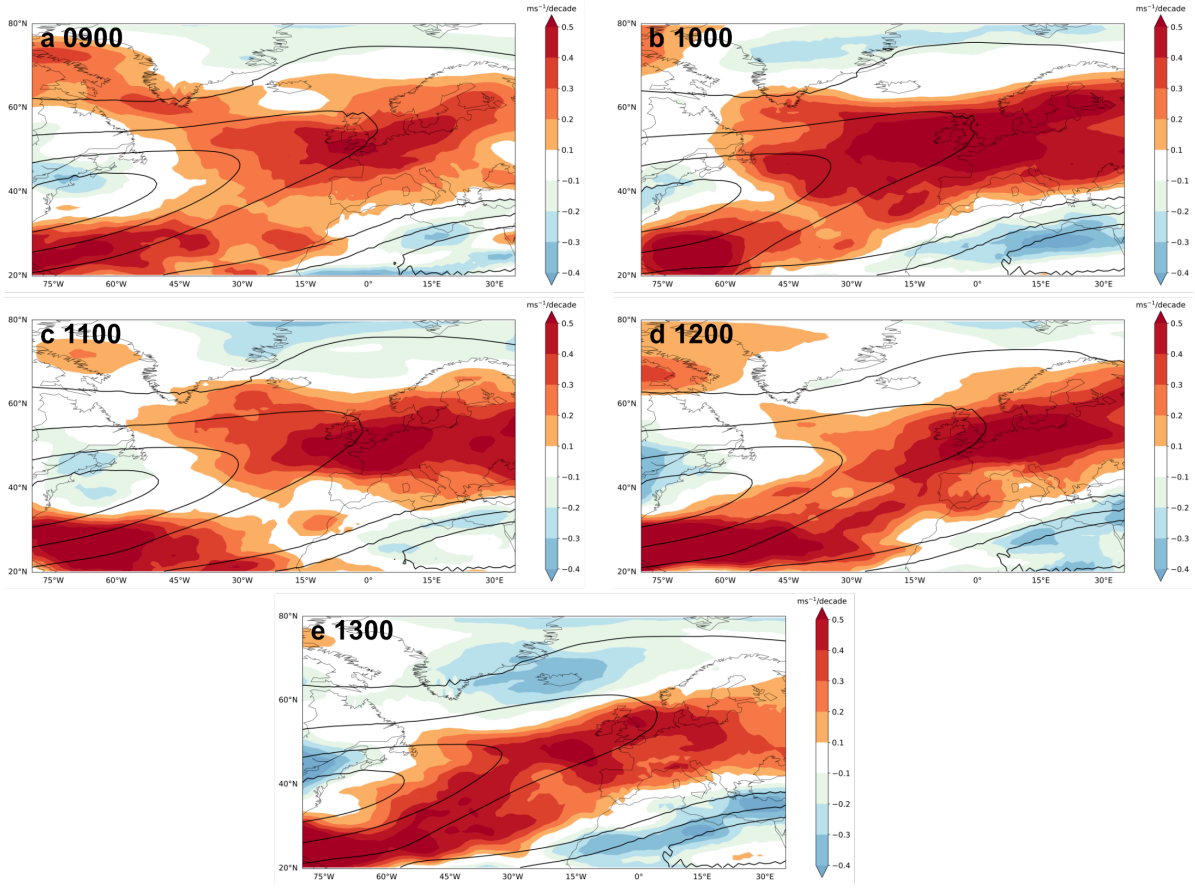

Figure 7: Forced trend in the geostrophic wind obtained from the LLAE (shading) and climatological mean (black contours, between 10 and 40 $\mathrm{m\,s^{-1}}$). Trends derived from the LLAE are the average of multiple training experiments in which the data of the respective ensemble member is not used in the training set. For example, panel a (member 0900) is the average trend resulting from LLAE trained on all combinations of the three ensemble members that do not include the member 0900.

ticular characteristics of individual trends are retained, for example the positive trend of member 0900 over the Labrador Sea, while the negative trend of member 1300 is removed. A possible explanation could be related to the selection of the domain, which is made to capture the main North Atlantic jet region and might not be optimal for other areas towards the vicinity of the boundaries in combination with low wind speed. The forced trend pattern however compares well to that obtained previously from the difference between the directly computed geostrophic wind from detrended and non-detrended geopotential (Fig. 3), indicating a satisfactory performance of the LLAE. Next, consideration is given to the full simulation period.

The forced trends in geostrophic wind for the whole simulation period (1980–2100, Fig. 8) also display a southwest-northeast band of increase in geostrophic wind, and a slight decrease over eastern North America and the Gulf Stream due to an equatorward shift of the jet in this sector. These trends are qualitative similar to those found for the present-day period but now much clearer and less nuanced, indicating that the long-term forced trend for the whole simulation period can be inferred already by the forced trend in the present-day period. The forced trend identified by the LLAE trained on geostrophic

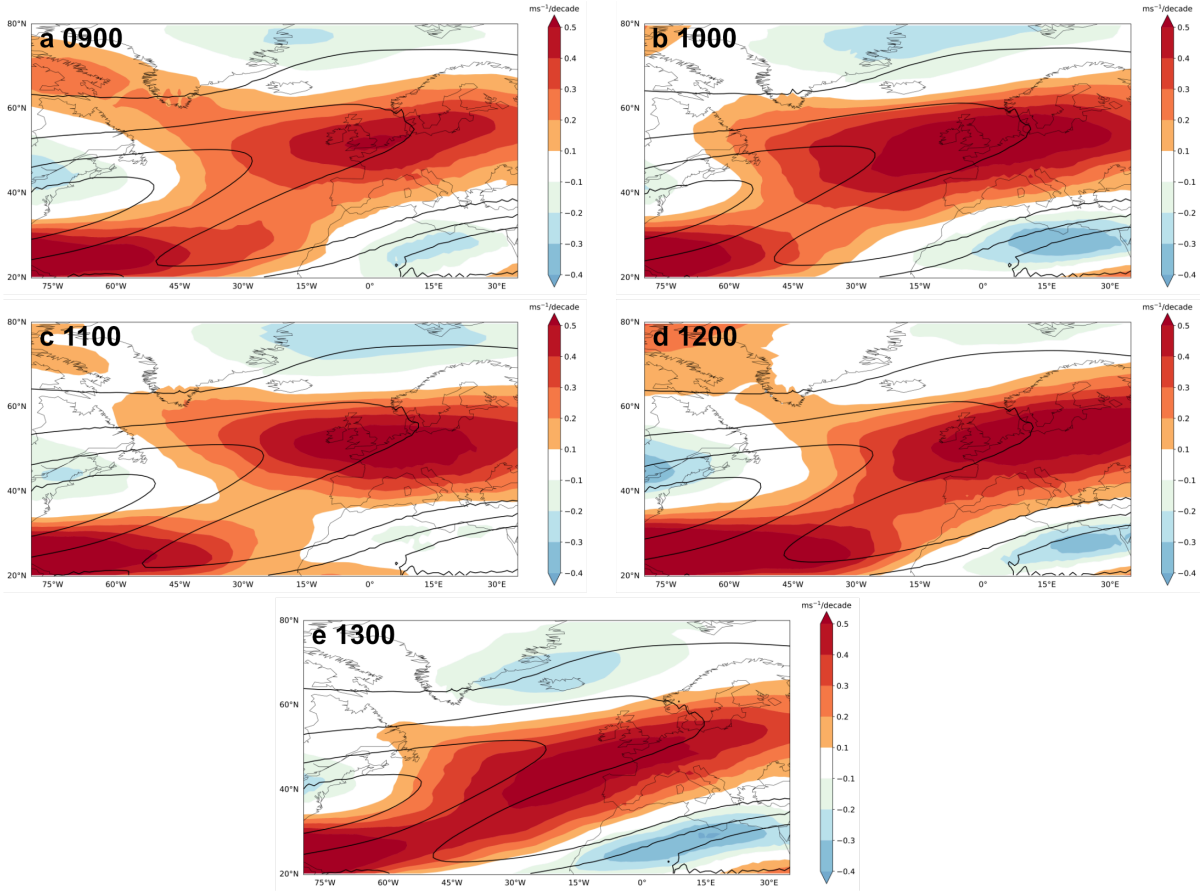

Figure 8: As in Fig. 7 for the period 1980–2100.

wind is also remarkably close to that obtained with the analytical geostrophic wind solution.

Forced trend in full wind Individual ensemble members display a large spread in the present-day wind trends, showing an intensification or an equatorward shift for some members while others indicate a poleward shift (Fig. 9 left column). This suggests that the different ensemble members may be sampling different modes of natural variability leading to a large ensemble spread. Nevertheless, a consistent trend pattern emerges based on the LLAE (Fig. 9 right column). The trends derived from the LLAE for the full wind show a similar pattern to the forced trends obtained using the geostrophic wind as target (variable Y) and are also similar to the forced trends obtained from direct computation of the geostrophic wind. This implies that forced trends in the full wind are mainly driven by forced changes in the geostrophic wind and thus by a modification of the pressure gradient force due to the external forcing.

The increase in zonal wind speed along a band with a southwest-northeast orientation is also a characteristic of the jet stream trend in ERA5 (not shown), which indicates that this change might result partly from external forcing. However, the positive trend in the reanalysis is located more poleward compared to the reanalysis. The fact that the LLAE is able to produce similar trends for the full wind for all ensemble members confirms that the method successfully removes the natural variability component that leads to different baroclinicity trends and consequently to contrasting jet stream

trends. However, a model bias with respect to ERA5 is still present and this cannot be removed with dynamical adjustment. The application of the LLAE to ERA5 data, although theoretically possible, is limited in practice by the lack of a sufficiently large dataset on which to train the model.

Throughout the simulation period, changes in the pressure gradient force are relevant, as the forced trends for the full zonal wind (right columns in Fig. 10) are similar to those obtained for the geostrophic wind and the present-day period. Another important aspect is that the consistency of the LLAE-based estimation across both periods. In the full simulation period the variability of the full trends among the ensemble members for the geostrophic wind is reduced (right column in Fig. 10), which suggests that the LLAE is capable of accurately quantifying the forced trend, even in the presence of larger member-to-member variability, as observed in the present-day period. This finding underscores the robustness of the LLAE in capturing the forced trend, despite significant inter-member variability, thereby enhancing the credibility of its results.

4 Conclusions

In this study, the Latent Linear Adjustment Autoencoder developed by Heinze-Deml et al. (2021) is used to disentangle trends in the winter North Atlantic jet forced by an increase in greenhouse gases from natural variability. Using detrended temperature and geopotential height, the model trained on data from an ensemble of five fully-coupled climate simulations can extract the forced trends in upper-level zonal wind. The efficacy of the model is particularly evident during the present-day period in which the original trends derived from each ensemble member display a large spread in the jet response. All ensemble members also share the main characteristics of the LLAE-derived trends for other periods with smaller ensemble spread, which highlights the robustness of the LLAE.

The North Atlantic jet stream displays a positive trend along a southwest-northeast oriented band, representing an equatorward shift of the jet in the entrance region and an extension towards Europe. These central elements are common for the different analyzed periods and are also found when analyzing forced trends directly from the difference between geostrophic wind computed from detrended[3] and non-detrended geopotential, which proves the adequate performance of the LLAE. Forced trends obtained from the LLAE using the full wind are similar to those resulting from geostrophic wind, suggesting that the main driver of the forced trend in the zonal wind speed is the change in the pressure gradient force, which determines the geostrophic component of the wind. The geostrophic wind is chosen to test the ability of the LLAE to learn how to predict the geostrophic wind because the geostrophic wind can be analytically compute from geopotential and an estimate of the forced trend for this field can be easily obtained.

Despite the satisfactory results, there are a few issues concerning the application of the LLAE to this problem. Forced trends in the geostrophic wind derived from the application of the LLAE exhibit some areas with smaller positive trends compared to the direct computation, especially over the ocean. This pattern is also seen in the forced trends for the full wind. Although the general forced trends from the LLAE agree with the direct computation, some caution is required in interpreting the small-scale structures produced by the LLAE, as these appear to be spurious. In addition, the performance of the model is strongly dependent on the variables used as input, and it is therefore necessary to ensure that the variables used as inputs explain a substantial part of

[3] We acknowledge that there are numerous methods for detrending a time series, and the exact method for removing the forcing trend is debated. For the purpose of this study, we assume that linear detrending is sufficient to demonstrate the ability of the LLAE to reconstruct the forcing trend in the zonal wind.

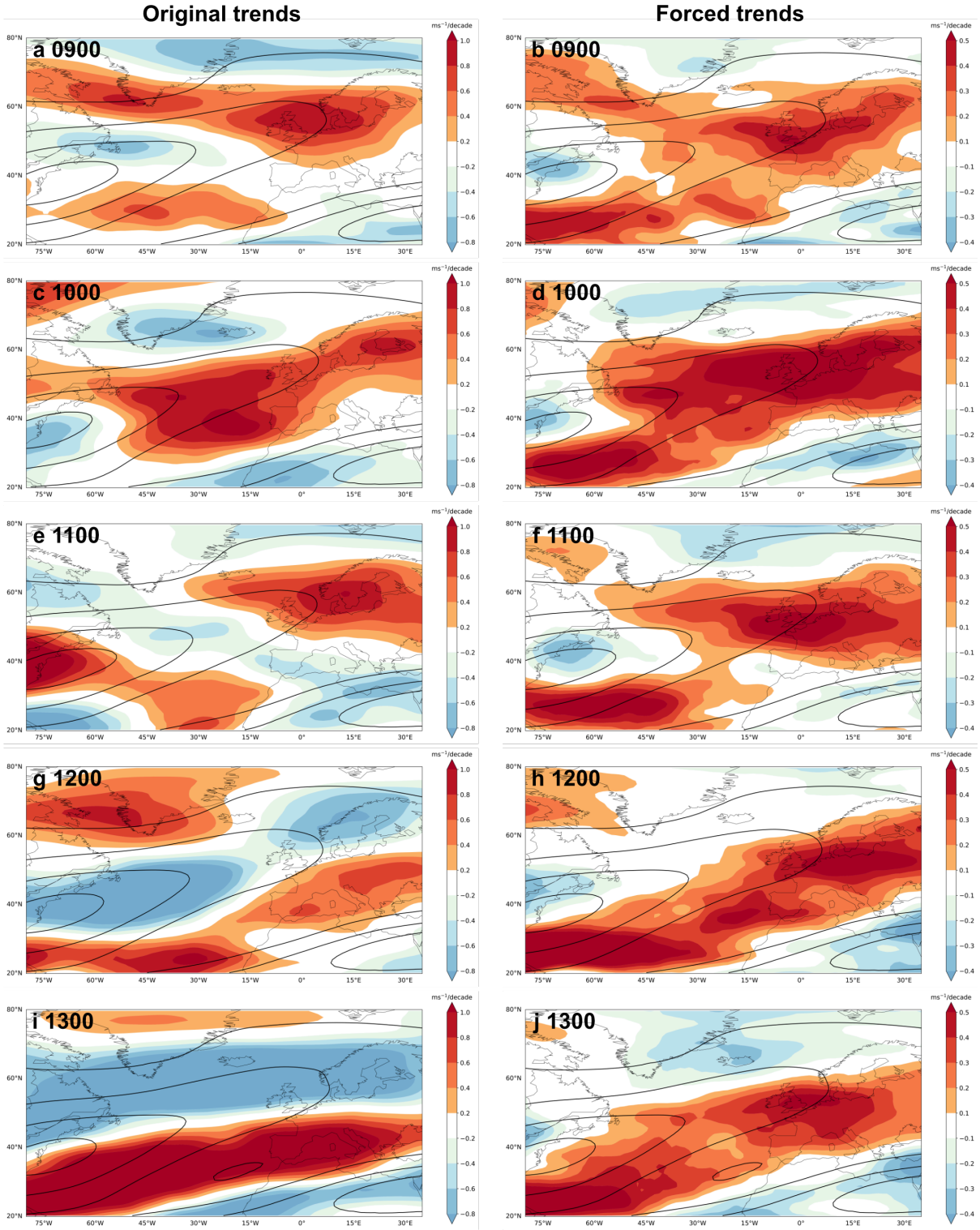

Figure 9: Left column: Linear trends of the zonal wind speed at 250 hPa for each ensemble member (shading) and climatological mean (black contours, between 10–40 $\mathrm{m\,s^{-1}}$) for the period 1980–2022. Right column: Linear trend of the residual between the zonal wind predictions provided by the LLAE and the original fields (shading) and climatological mean (black contours, between 10–40 $\mathrm{m\,s^{-1}}$). Trends derived from the LLAE are the average of multiple training experiments in which the data of the respective ensemble member is not used in the training set. Note that the range of the color scale for the original trends is twice the range of the forced trends.

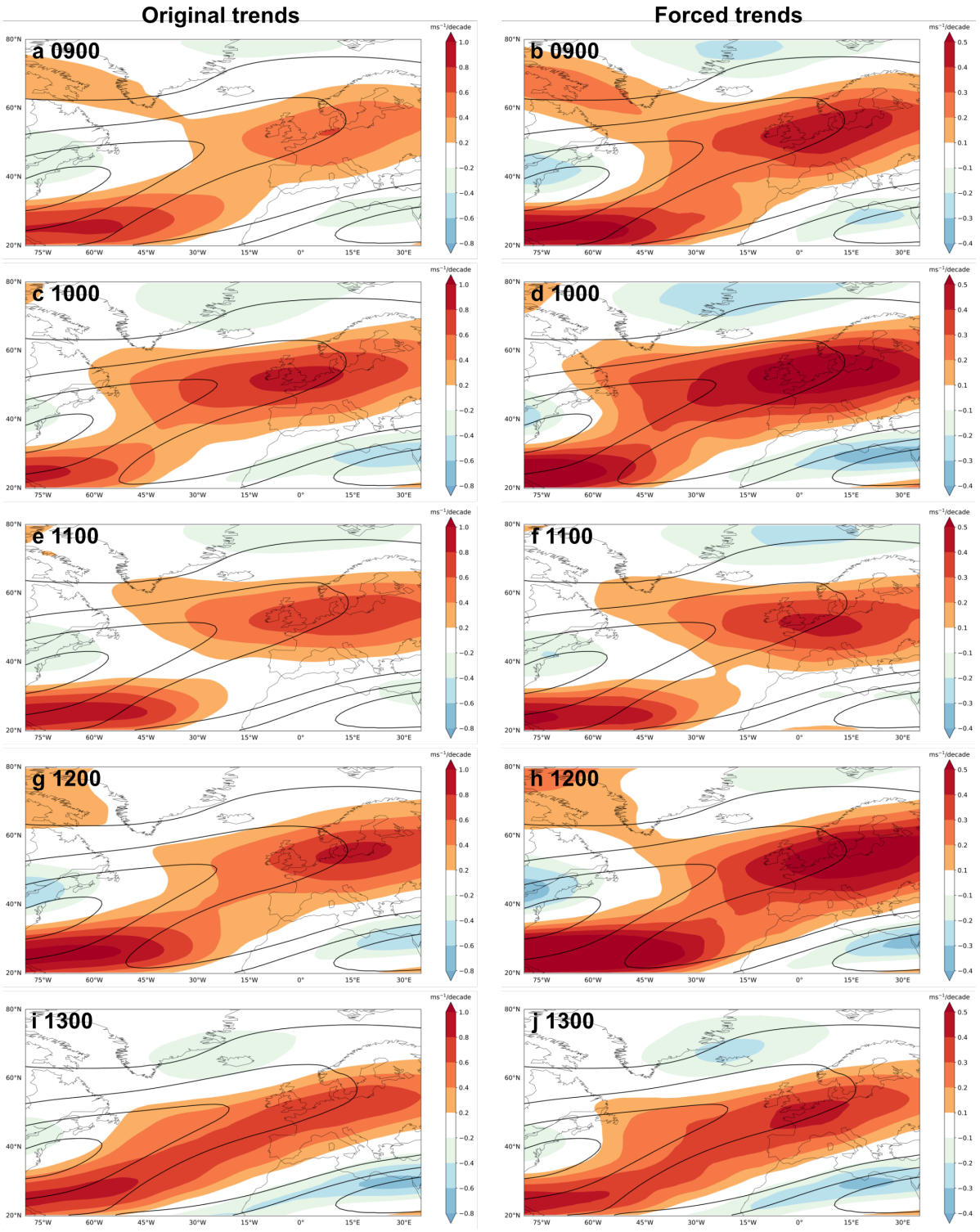

Figure 10: As in Fig. 9, but for the period 1980–2100.

the variance of the variable of interest in order to produce meaningful residuals that contain the part of the trend generated by external forcing and not additional terms related to an inadequate choice of input variables. Although the model can be applied in theory to a small ensemble or to observational data, it requires a sufficiently large dataset to avoid overfitting. This aspect limits in practice the use of the LLAE to a certain applications, such as dynamical adjustment of observational or reanalysis data.

5 Open Research

The Latent Linear Adjustment Autoencoder model is free and open source. It is distributed under the MIT software license which allows unrestricted use. The source code is available at the following GitHub repository: `https://github.com/christinaheinze/latent-linear-adjustment-autoencoders`. CESM simulation data will be stored at the Institute of Atmospheric and Climate Science, ETH Zurich, for at least 10 years and are available on request.

Acknowledgments
This work was supported by the Swiss National Science Foundation through project grant Nr. 204181.

References

Bengtsson, L., & Hodges, K. I. (2019). Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability? *Climate Dynamics*, *52*, 3553–3573. doi: 10.1007/s00382-018-4343-8

Blackport, R., & Fyfe, J. C. (2022). Climate models fail to capture strengthening wintertime North Atlantic jet and impacts on Europe. *Science Advances*, *8*. doi: 10.1126/sciadv.abn3112

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., & Schanen, D. P. (2018). The path to CAM6: coupled simulations with CAM5.4 and CAM5.5. *Geoscientific Model Development*, *11*, 235–255. doi: 10.5194/gmd-11-235-2018

Connolly, C., Barnes, E., Hassanzadeh, P., & Pritchard, M. (2023). Using neural networks to learn the jet stream forced response from natural variability. *Artificial Intelligence for the Earth Systems*, *2*, 1-15. doi: 10.1175/AIES-D-22-0094.1

Dai, A., & Bloecker, C. E. (2019). Impacts of internal variability on temperature and precipitation trends in large ensemble simulations by two climate models. *Climate Dynamics*, *52*, 289–306. doi: 10.1007/s00382-018-4132-4

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., ... Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001916. doi: 10.1002/qj.2378

Deser, C., Hurrell, J. W., & Phillips, A. S. (2017). The role of the North Atlantic Oscillation in European climate projections. *Climate Dynamics*, *49*, 3141–3157. doi: 10.1007/s00382-016-3502-z

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, *38*, 527–546. doi: 10.1007/s00382-010-0977-x

Deser, C., & Phillips, A. S. (2023). A range of outcomes: the combined effects of internal variability and anthropogenic forcing on regional climate trends over Europe. *Nonlinear Processes in Geophysics*, *30*, 63—84. doi: 10.5194/npg-30-63-2023

Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, *29*, 2237–2258. doi: 10.1175/JCLI-D-15-0304.1

de Vries, I. E., Sippel, A. G., S. Pendergrass, & Knutti, R. (2023). Robust global detection of forced changes in mean and extreme precipitation despite observational disagreement on the magnitude of change. *Earth System Dynamics*, *14*, 81–100. doi: 10.5194/esd-14-81-2023

Guan, X., Huang, J., Guo, R., & Lin, P. (2015). The role of dynamically induced variability in the recent warming trend slowdown over the Northern Hemisphere. *Scientific Reports*, *5*, 12669. doi: 10.1038/srep12669

Guo, R., Deser, C., & Terray, L. (2019). Human influence on winter precipitation trends (1921–2015) over North America and Eurasia revealed by dynamical adjustment. *Geophysical Research Letters*, *46*, 3426–3434. doi: 10.1029/2018GL081316

Heinze-Deml, C., Sippel, S., Pendergrass, A. G., Lehner, F., & Meinshausen, N. (2021). Latent Linear Adjustment Autoencoder v1.0: a novel method for estimating and emulating dynamic precipitation at high resolution. *Geoscinetific Model Development*, *14*, 4977–4999. doi: 10.5194/gmd-14-4977-2021

Hunke, E., Lipscomb, W., Turner, A., & Jeffery, S., N. ans Elliott. (2015). *ICE: The Los Alamos Sea Ice Model documentation and software user's manual version 5.1 LA-CC-06-012* (Tech. Rep.). Santa Fe, NM, USA: Los Alamos National Laboratory.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*, 676–679. doi: 10.1126/science.269.5224.6

Kingma, D. P., & Ba, J. (2015). A method for stochastic optimization. In *3rd international conference on learning representations, san diego, ca, usa.*

Kingma, D. P., & Welling, M. (2021). An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, *12*, 307–392. doi: 10.1561/2200000056

Lawrence, D., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., ... Zeng, X. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*, 4245-4287. doi: 10.1029/2018MS001583

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. doi: 10.1038/nature14539

Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of "time of emergence" of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, *30*, 7739–7756. doi: 10.1175/JCLI-D-16-0792.1

Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, L. R. (2013). A physically based runoff routing model for land surface and Earth system models. *Journal of Hydrometeorology*, *14*, 808–828. doi: 10.1175/JHM-D-12-015.1

McKenna, C. M., & Maycock, A. C. (2021). Sources of uncertainty in multimodel large ensemble projections of the winter North Atlantic Oscillation. *Geophysical Research Letters*, *48*, e2021GL093258. doi: 10.1029/2021GL093258

Novak, L., Ambaum, M. H. P., & Tailleux, R. (2015). The life cycle of the North Atlantic storm track. *Journal of the Atmospheric Sciences*, *72*, 821–833. doi: 10.1175/JAS-D-14-0082.1

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482. doi: 10.5194/gmd-9-3461-2016

Schemm, S., Rivière, G., Ciasto, L. M., & Li, C. (2018). Extratropical cyclogenesis changes in connection with tropospheric ENSO teleconnections to the North Atlantic: Role of stationary and transient waves. *Journal of the Atmospheric Sciences*, *75*, 3943–3964. doi: 10.1175/JAS-D-17-0340.1

Simpson, I. R., Shaw, T. A., & Seager, R. (2014). A diagnosis of the seasonally and longitudinally varying midlatitude circulation response to global warming. *Journal of the Atmospheric Sciences*, *71*, 2489—2515. doi: 10.1175/JAS-D-13-0325.1

Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., & Knutti, R. (2019). Human influence on winter precipitation trends (1921–2015) over North America and Eurasia revealed by dynamical adjustment. *Journal of Climate*, *32*, 5677–5699. doi: 10.1175/JCLI-D-18-0882.1

Smoliak, B. V., Wallace, J. M., Lin, P., & Fu, Q. (2015). Dynamical adjustment of the Northern Hemisphere surface air temperature field: Methodology and application to observations. *Journal of Climate*, *28*, 1613–1629. doi: 10.1175/JCLI-D-14-00111.1

Terray, L. (2021). A dynamical adjustment perspective on extreme event attribution. *Weather and Climate Dynamics*, *2*, 971–989. doi: 10.5194/wcd-2-971-2021

Yadav, P., & Straus, D. M. (2017). Circulation response to fast and slow MJO episodes. *Monthly Weather Review*, *145*, 1577–1596. doi: 10.1175/MWR-D-16-0352.1