Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL081725

Are Midtwentieth Century Forced Changes in North Atlantic Hurricane Potential Intensity Detectable?

L. Trenary[1,2], T. DelSole[1,2], S. J. Camargo[3], and M. K. Tippett[4]

[1]Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, VA, USA, [2]Center for Ocean-Land-Atmosphere Studies, Fairfax, VA, USA, [3]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA, [4]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

Key Points:
- Forced changes in the basin mean that tropical North Atlantic hurricane potential intensity are detected in CMIP5 models
- The forced response is weak and differs across models, with some models simulating an increase and others a decrease
- Model differences and model reanalysis inconsistencies make attribution of observed changes in North Atlantic potential intensity nonviable

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2

Correspondence to:
L. Trenary,
ltrenary@gmu.edu

Citation:
Trenary, L., DelSole, T., Camargo, S. J., & Tippett, M. K. (2019). Are midtwentieth century forced changes in North Atlantic hurricane potential intensity detectable? *Geophysical Research Letters*, *46*, 3378–3386. https://doi.org/10.1029/2018GL081725

Received 20 DEC 2018
Accepted 21 FEB 2019
Accepted article online 4 MAR 2019
Published online 19 MAR 2019

Abstract The impact of anthropogenic forcings on tropical North Atlantic hurricane potential intensity (PI) is evaluated in Climate Model Intercomparison Project 5 models for the period 1958–2005. Eleven models are examined, but only seven models have a forced response that is distinguishable from internal variability. The use of discriminant analysis to optimize detectability does not yield a clear, common climate change signal. Of the seven models with a significant response, one has a negative linear trend while two have a positive linear trend. The trend in PI is not even consistent among reanalyses, although this difference is not statistically significant because of large uncertainties. Furthermore, estimates of PI internal variability have significantly different variances among different reanalysis products. These disagreements between models, reanalysis products, and between models and reanalyses, in conjunction with relatively large uncertainties, highlight the difficulty of detecting and attributing observed changes in North Atlantic hurricane potential intensity.

Plain Language Summary Observed temperature has been steadily increasing over the last century and much of this warming can be attributed to greenhouse gas emissions. Theoretically, the maximum intensity (or potential intensity) a hurricane can achieve depends strongly upon sea surface temperature, with warmer temperatures producing stronger storms. From this perspective, we might expect that the warming surface temperatures are driving observable changes in hurricane intensity. To this end, we analyze climate model experiments to determine if the observed changes in North Atlantic hurricane intensity can be attributed to human-related emissions over the period 1958–2005. Of the 11 models analyzed, we find that only seven predict that hurricane potential intensity has changed in response to greenhouse gas and aerosol emissions. The change in potential intensity differs across models, with one model predicting a decreasing trend in North Atlantic hurricane potential intensity, while two models predict an increasing trend in potential intensity. Different reanalysis data sets are likewise inconsistent. These results indicate that currently we cannot attribute changes in North Atlantic hurricane intensity to human-related forcings. It is possible that as greenhouse gas concentrations continue to increase, an unequivocal forced response in North Atlantic potential intensity may emerge in the future.

1. Introduction

Hurricanes making landfall in the United States are responsible for catastrophic losses in life and property, with nearly 3,500 fatalities since 1980 and an accumulated cost of $850.5 billion dollars in damages (Smith, 2018). With nearly half of the U.S. population living along the coasts, and the number of residents projected to increase, the potential risk of U.S. coastal communities to tropical cyclones is expected to grow. Given these increased vulnerabilities, it is important to understand how climate change may amplify risk in these densely populated regions.

Greenhouse gas (GHG) forced changes in sea surface temperatures (SSTs) may lead to a significant increase in hurricane intensity. Emanuel (1987) estimated that a doubling of carbon dioxide would lead to a 40–50% increase in hurricane potential intensity. High-resolution projections likewise find an increase in hurricane intensity with increasing GHG concentrations (Bengtsson et al., 2007; Knutson & Tuleya, 2004; Knutson et al., 2010; Maloney et al., 2013; Walsh et al., 2015). More than half of the observed increase in global mean surface temperature between 1951 and 2010 is very likely due to increases in GHG concentrations

(Bindoff et al., 2013). We might expect that these changes in surface temperature are reflected in observable changes in hurricane intensity.

Storm-specific observations document a strengthening of North Atlantic hurricanes in recent decades. Satellite records capture a significant shift in the distribution of North Atlantic hurricane intensity toward stronger storms since the 1980s (Elsner et al., 2008; Kossin et al., 2013). Likewise, best track observations of the power dissipation index of North Atlantic hurricanes, an aggregate measure of hurricane frequency, intensity, and duration, has steadily increased since the 1970s (Emanuel, 2005). While large uncertainties in best track records have resulted in considerable debate about the validity of observed trends in hurricane intensity (Klotzbach & Landsea, 2015; Webster et al., 2005), the advent of temporally homogenous satellite data sets has allowed for more reliable estimates of hurricane intensity over recent decades (e.g., Kossin et al., 2013).

Trends have also been identified in the large-scale proxy for hurricane strength known as the potential intensity (PI). The PI provides an estimate of the theoretical maximum intensity a hurricane can achieve given the surface temperature and vertical profiles of temperature and moisture. Although there are well known caveats in the PI theory (Smith et al., 2008), an alternative measure has not been proposed. Both observational and modeling studies document a relation between PI and average hurricane intensity, where increases in PI correspond with stronger cyclones (Emanuel, 2000; Kossin & Camargo, 2009; Wing et al., 2007). Consistent with the storm-based measurements, positive trends in the basin mean tropical North Atlantic PI have been identified in a variety of reanalysis products and in radiosonde data over recent decades (Emanuel et al., 2013; Vecchi et al., 2013; Wing et al., 2015). In contrast, Kossin (2015) found that the sign of the trend in storm-local North Atlantic PI is dependent on the reanalysis product being analyzed. Multiple lines of evidence suggest that the intensity of North Atlantic hurricanes has changed over recent decades, but the cause of this change remains a subject of debate.

The frequency of major hurricanes in the North Atlantic is strongly linked to the local SST changes, where strong storms are more likely when SSTs are warm (Yan et al., 2017). Fluctuations in observed North Atlantic SST are the result of external forcing and internal variability, though it is unclear which contributes most to changes in local SST. Analyzing a suite of Climate Model Intercomparison Project (CMIP5) experiments, Ting et al. (2015) argue that the impact of GHG and aerosol forcings on North Atlantic hurricane potential intensity cancel and attribute trends in PI over the recent decades to the Atlantic multidecadal variability. The cancelation of the aerosols and GHG forcing on PI globally is discussed in Sobel et al. (2016, 2018).

The impacts of GHG and aerosol forcing on hurricane intensity cannot be determined from observations alone. To this end, we analyze climate model experiments to determine if the observed changes in North Atlantic PI can be attributed to human-related forcings by comparing estimates of the forced response from models with reanalyses. Our analysis will focus on the PI, because it serves as a useful proxy for the distribution of hurricane intensity in global climate models (e.g., Camargo et al., 2013; Polvani et al., 2016; Sobel et al., 2016; Ting et al., 2015), which lack sufficient spatial resolution to explicitly simulate the most intense hurricanes (Camargo & Wing, 2016). Formal attribution and detection analysis of the single forced runs reveal a collinearity in PI to GHG and aerosol forcings. Therefore, rather than attempting to attribute changes in PI to GHG and aerosol forcings separately, we evaluate the impacts of the combined forcings. Previously, Ting et al. (2015) examined the impacts of external forcing and internal variability on SST-related changes on North Atlantic PI in a suite of CMIP5 models. Unlike Ting et al. (2015), which was a multimodel study, we quantify the combined impacts of GHG and aerosol forcings on PI variability in individual CMIP5 models. Lastly, the consistency of the forced response is assessed by a cross-model comparison and verified against different reanalyses.

2. Data and Methods

We analyze the historical simulations, wherein external forcings change according to observed estimates, and the preindustrial controls in which forcings are fixed, for the 11 climate models listed in Table S1 in the supporting information. The anthropogenically forced response for each model is estimated by averaging across all available ensemble members for the historical simulations (see Table S1). One ensemble member from the preindustrial control is used to estimate the internal variability. We consider only the period 1958–2005, to allow for comparison with reanalysis products. The analysis was repeated for the period 1979–2005 and the results are consistent; however, the number of models with a detectable trend is smaller

than that for the 1958–2005 period. Accordingly, we limit our analysis to the longer period to allow for the inclusion of more models. A 240-year segment is used from each model preindustrial control, of which 192 years are used to estimate internal variability, the remaining 48 years are used in hypothesis testing to match the observational data. The model with the least number of years limits the length of the preindustrial control run used. For both the historical and preindustrial control experiments, monthly fields of SST and vertical profiles of temperature and moisture are first interpolated to a $2° \times 2°$ grid and then used to estimate PI according to Bister and Emanuel (1998). We also analyzed the single forced (GHG only and aerosol only) simulations of these same models, but those results are not reported here. The remainder of our analysis focuses on the anomalies of the average June–November tropical North Atlantic PI, where anomalies are estimated with respect to overall mean for the period of interest. A subset of the models selected for this study were also analyzed in Ting et al. (2015) and Sobel et al. (2016). For a more complete description of the models, please refer to Taylor et al. (2012).

Historical changes in the modeled PI are compared with estimates obtained from four reanalyses products—the European Centre for Medium-Range Weather Forecasts (ECMWF) twentieth century (ERA-20C; Poli et al., 2016), the ECMWF-Interim (ERA-Int; Dee et al., 2011), the 55-year reanalysis produced by the Japanese Meteorological Agency (JRA-55; Kobayashi et al., 2015), and the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) reanalysis (Kalnay et al., 1996). We find that the low-frequency PI variability in ERA-Int is consistent with ERA-20C and ERA-20C has a longer record; consequently, most of our discussion is limited to results based on ERA-20C. The PI for reanalyses is computed using the same scripts as in the models.

Impacts of anthropogenic climate change are likely to be seen at the largest spatial scales. While most studies evaluating long-term changes in PI focus on the basin mean, it is possible that more complex spatial structures are needed to capture the impacts of external forcing on PI. For this reason, we isolate the large-scale variations by projecting PI anomalies from the historical and preindustrial control runs onto the leading eigenvectors of the Laplacian operator over the tropical North Atlantic (0–20°N, see Figures S1a–S1d). This region was selected because it encompasses the portions of tropical Atlantic where most tropical cyclones form. The eigenvectors are ordered by decreasing spatial scale, where the first vector gives the basin mean, followed by dipoles, tripoles, quadrupoles, and so forth (DelSole & Tippett, 2015). The projection produces a set of time series that represent the temporal evolution of each pattern. The spatial patterns are data independent and therefore serve as a useful basis set when comparing data sets since the time series for each Laplacian always represents a common spatial structure. In section 3, we will show that forced changes in tropical North Atlantic PI are primarily captured by the basin mean (i.e., the first Laplacian).

Discriminant analysis is used to isolate the externally forced changes in the PI. The method is a standard optimization technique that determines the pattern that maximizes the ratio of variance in the forced run to the variance in the preindustrial control (DelSole & Tippett, 2007; Schneider & Held, 2001). The maximization leads to an eigenvalue problem

$$\Sigma_{\mathbf{f}}\mathbf{q} = \lambda\Sigma_{\mathbf{c}}\mathbf{q}, \tag{1}$$

where $\mathbf{\Sigma_f}$ is the covariance matrix estimated from the Laplacian time series of the historical run ensemble mean, and $\mathbf{\Sigma_c}$ is covariance matrix estimated from the Laplacian time series of a single ensemble from the preindustrial control. The term $\mathbf{q}$ is the projection vector, and λ is the eigenvalue giving the variance ratio. The eigenvectors are ordered by decreasing eigenvalues, such that the first maximizes the variance ratio, and the second maximizes the ratio subject to being uncorrelated to the first and so on. Prior to applying discriminant analysis, the time series obtained from the historical run ensemble average is smoothed by regressing onto the first three Legendre polynomials (i.e., a trend, a parabola, and cubic). The smoothing helps isolate the forced response by filtering any remaining internal variability. Moreover, the coefficients recovered from the regression provide a means for quantifying differences in the forced response across data sets in a statistically rigorous way.

The significance of the optimized variance ratio (λ) is determined by repeating the optimization using subsamples of the preindustrial control run. Specifically, a continuous 48-year segment from the preindustrial control is selected and the Laplacian time series are smoothed and a covariance matrix constructed, consistent with our treatment of the historical run data described above. The remaining 192 years are used to estimate the covariance matrix of the unsmoothed Laplacian time series (i.e., $\mathbf{\Sigma_c}$). The optimization is then

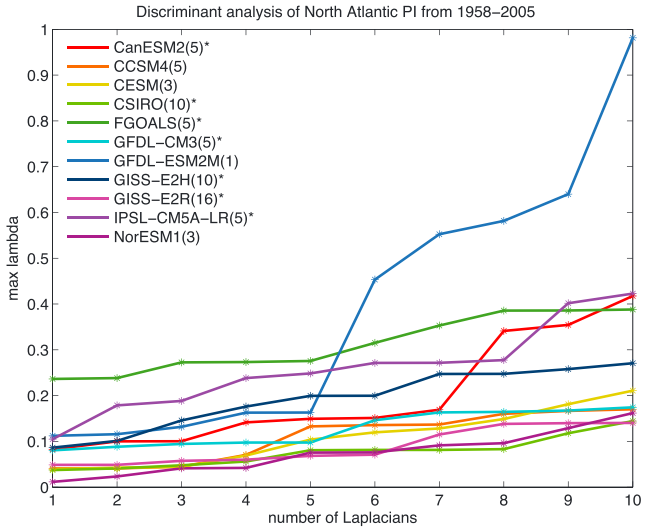

Figure 1. Maximized signal-to-noise ratio of mean June–November anomalous PI for a subset of CMIP5 models as a function of the number of Laplacian eigenfunctions (the more eigenfunctions, the smaller the resolved spatial scale). The climate model names in the legend that are followed by an asterisk identify models that capture a statistically significant change in basin mean North Atlantic PI. The number listed parenthetically next to the model name gives the ensemble size. Significance is estimated by repeating the optimization using subsamples of the preindustrial control run to build up an empirical distribution of the maximum λ. The 99th percentile of the empirical distribution serves as the threshold for detectability.

repeated for each paired 48 and 192-year segments in the preindustrial control simulation, yielding ~192 estimates of the eigenvalues (λ). These results are used to construct an empirical distribution of the leading λ consistent with the hypothesis of pure internal variability, from which the 99th percentile defines the significance threshold or threshold for detectability. The significance is confirmed using three different models of internal variability: white noise process and first- (AR1) and second- (AR2) order autoregressive processes. For each of these alternative hypotheses, synthetic data with properties consistent with the underlying null hypothesis are generated and the optimization is repeated 5,000 times to build up a distribution of the leading eigenvalues (λ) and the significance threshold for detectability is estimated as the 99th percentile of the distribution. All four significance tests take into account the reduction in variance that results from ensemble averaging by dividing the recovered eigenvalues by the appropriate ensemble size. Consequently, the significance threshold for each model is different, since each model has a different number of ensemble members.

3. Results

The forced response of PI to individual forcings was initially estimated by applying discriminant analysis to the single forced runs. A formal attribution and detection analysis, however, revealed that the two responses were highly collinear and difficult to separate (not shown). This is consistent with the findings of Ting et al. (2015) and Sobel et al. (2016), where collinearity was inferred from the difference in GHG and aerosol forcings estimated from multimodel/multiensemble averages. Therefore, we estimate the *total* forced response by applying discriminant analysis to the historical simulations that contain both GHG and aerosol forcings. To further improve estimates of the forced response in PI, the ensemble mean from the historical runs is used in the discriminant analysis. We begin our analysis by examining which spatial structures are needed to fully characterize the forced response in North Atlantic PI. To do so the discriminant analysis is repeated for each model using an increasing number of Laplacians time series. The resulting signal-to-noise ratios are shown in Figure 1, each curve gives the values of the leading discriminant (λ; vertical axis) as a function of the number of included patterns (shown on the horizontal axis). We limit our analysis to 10 Laplacian because the inclusion of more spatial structures does not impact our conclusions. Of the 11 models analyzed, seven models (indicated by the asterisk in the legend of Figure 1) detect a forced response in PI when only one Laplacian eigenvector is used. The significance is evaluated with respect to the subsampled preindustrial control data; however, the statistical significance is insensitive to the underlying null hypothesis

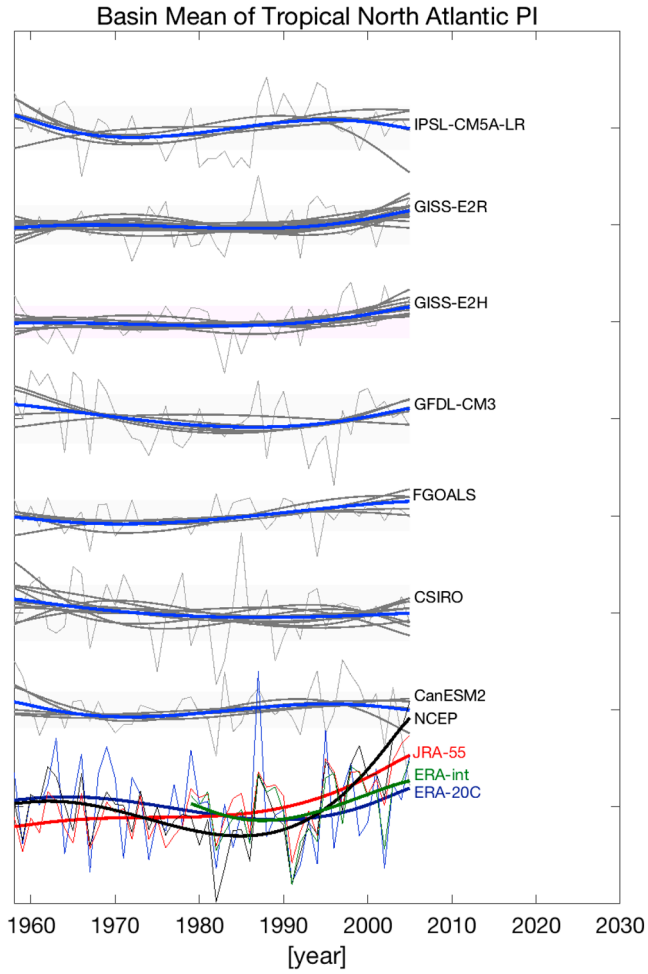

Figure 2. Time series of the anomalous mean June–November tropical North Atlantic PI estimated for various climate models and reanalysis products. The forced response is shown as the thick curves and is estimated by regressing the time series of the basin mean PI onto the first three Legendre polynomials. The thick blue curves give the forced response estimated from the ensemble mean. The thick gray curves give the forced response for individual ensemble members. The thin gray lines represent the unsmoothed basin mean PI for a single ensemble. The gray bounding boxes represent the 1 standard deviation range estimated with respect to 192-year unsmoothed data from the model's control run. Estimates of the PI obtained from JRA-55 (red), ERA-20C (blue), ERA-Int (green), and NCEP (black), where the thick curves give an estimate of the forced response and the thin curves represent the unsmoothed basin mean PI.

(not shown). The inclusion of higher-order Laplacians does increase the signal-to-noise ratio (λ); however, the time series describing the signal (the leading discriminant time series) generally changes very little beyond the first Laplacian (see Figure S2), indicating that the detectable changes in PI are primarily associated with changes in the basin mean. Aside from CanESM2 and to a lesser extent IPSL-CM5A-LR, the inclusion of more complex spatial structures does not change the time series of the leading discriminant. As for CanEMS2 and IPSL, differences are only evident when a large number of Laplacian structures are included. Four models (CCSM4, CESM, GFDL-ESM2M, and NorESM1) show no significant change in the basin mean PI. The inclusion of more patterns does lead to significant signal-to-noise ratios in CCSM4 and GFDL-ESM2M, but not CESM and NorESM1 (not shown). Since this is a model-dependent result and there is no established criterion for preferring one model over another, we limit our analysis to the basin mean (first Laplacian), and therefore exclude CCSM4, CESM, GFDL-ESM2M, and NorESM1 from our analysis, because none of these models yields a detectable signal when only one Laplacian eigenvector is used.

Note that the spatial pattern is exactly the same for all models because we are examining time series from the first Laplacian eigenvector. Thus, the forced responses among the models can only differ in their time series. For each model, the variations of the basin mean tropical North Atlantic PI anomalies, shown in

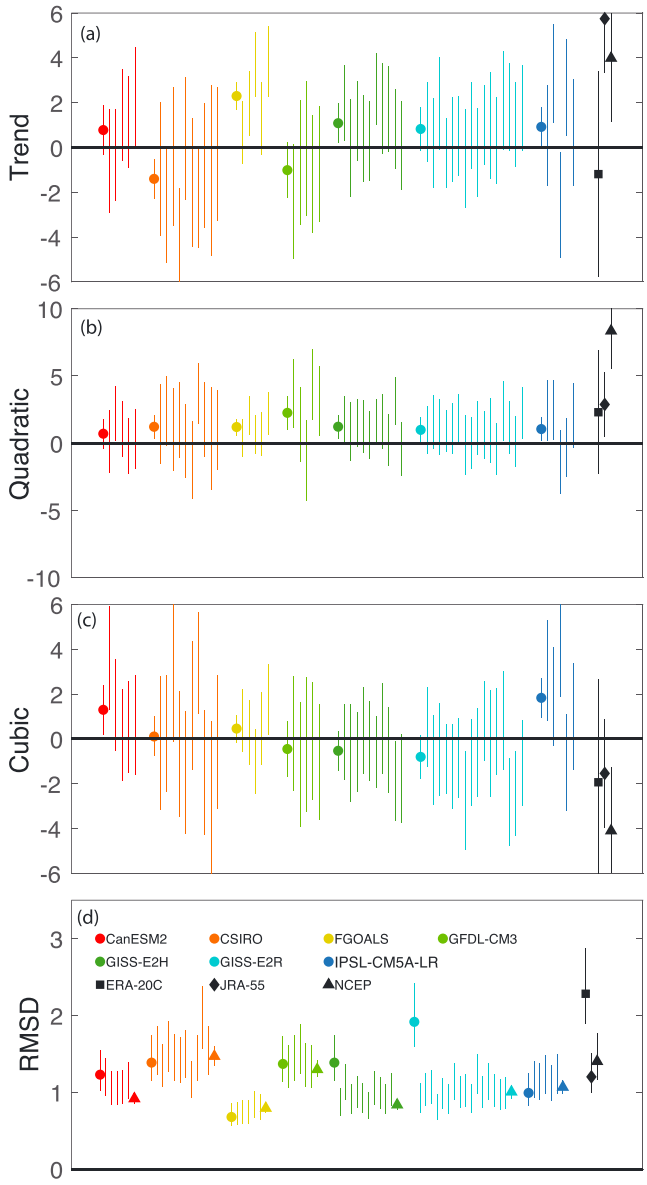

Figure 3. Coefficients estimated from the regression of the time series of the anomalous mean June–November tropical North Atlantic PI onto the (a) trend, (b) parabolic, and (c) cubic Legendre polynomials. Regression coefficients for the model's ensemble mean are represented by colored circles. The colored intervals give the confidence intervals for estimates obtained from individual ensemble members for the specified model. Coefficients for reanalysis products are given by the black symbols. The bars show the 95% confidence intervals. The corresponding root-mean-square difference (RMSD) of the regression residual is shown in (d). The colored triangle represents the standard deviation of the first Laplacian time series estimated from 192 years of preindustrial control data. RMSD for the ensemble mean estimates are multiplied by the square root of the ensemble size to account for the reduction in variance associated with ensemble averaging.

Figure 2, are largely confined to be less than 1 standard deviation of the preindustrial control (gray-shaded region), indicating that the detected changes correspond to a low signal-to-noise ratio. Comparing the temporal evolution of the low-frequency ensemble mean PI (thick blue curve in Figure 2), it appears that the response differs across the models. Comparing the models and reanalyses (curves, bottom of Figure 2), it is clear that the variations of low-frequency PI differ across models, between models and reanalyses, and among the reanalysis products. Because the low-frequency estimate of PI variability in ERA-Int is consistent with ERA-20C, the remainder of our analysis will focus on ERA-20C since it provides a longer record of PI variability.

To quantify differences in the evolution of the time series, we compare the regression coefficients for the three Legendre polynomials (i.e., trend, quadratic, cubic) for the ensemble mean and individual ensemble members. From Figure 3, we see that the coefficient estimates for individual ensemble members are largely indistinguishable from zero and are generally consistent with each other and the ensemble mean. The few exceptions are no greater than what might be expected by chance. Ensemble averaging reduces uncertainty in coefficients estimates. Moreover, we find that when the ensemble mean is used, at least one coefficient among the three Legendre polynomials is distinguishable from zero. In contrast, none of the four models that were excluded had a significant Legendre coefficient (not shown), further supporting our conclusion that these models have no detectable change in PI. For the seven models that do detect a change in PI, the nonzero coefficients differ across the models. Differences across models and reanalyses are most evident in the sign of the linear trend term: three models (FGOALS, GISS-E2R, and IPSL-CM5A-LR) capture a statistically significant positive trend, indicating an increase in PI, and one model (CSIRO) captures a statistically significant negative trend or decrease in PI. The trends for the other four models are statistically indistinguishable from zero. Reanalyses similarly differ with respect the trend sign: JRA-55 and NCEP both capture a statistically significant positive trend or increase in PI, while the trend in ERA-20C is negative and indistinguishable from zero. The positive trends in tropical North Atlantic PI in various reanalysis products were also reported in Vecchi et al. (2013), Emanuel et al. (2013), and Wing et al. (2015). The greatest consistency between models and reanalysis is related to the sign of the quadratic term, while not always significant, the coefficients are positive. Differences are also evident in the sign of the cubic term, with two models (CanESM2 and IPSL-CM5A-LR) capturing a statistically significant positive coefficient. The coefficients for the other models varies in sign and are statistically insignificant. Notably, the quadratic trend for NCEP is much larger than model estimates.

The above results clearly show that models differ in their forced responses and that reanalysis products differ in the low-frequency variability. We now evaluate model reanalysis consistency of internal variability. If the forced response is sufficiently captured by the Legendre polynomials, then the residuals should provide independent estimates of the unforced PI variability. More precisely, the root-mean-square difference (RMSD) from the regression residual for the ensemble mean (circles in Figure 3d) and individual ensemble (confidence intervals, Figure 3d) should be consistent with the standard deviation of the PI from the preindustrial control runs (colored triangle in Figure 3d). With respect to individual models, generally, there is a consistency between the different estimates of internal variability (i.e., overlap in the confidence intervals). There are a few exceptions, the ensemble mean estimates of the RMSD for GISS-E2R and GISS-E2H, are inconsistent with individual ensemble estimates and the model's internal variability. The residuals from both these models are dominated by a common signal that appears to be related to volcanic eruptions that are not captured by the polynomial fit (not shown). Generally, model estimates of internal variability are consistent across models. For reanalysis, we find that the magnitude of the RMSD is substantially larger for estimates obtained from ERA-20C compared to JRA-55 and NCEP. This suggests that there are high-frequency variations in ERA-20C PI that are not present in the other reanalyses or the models. A formal F test confirmed that the variance of the first Laplacian time series obtained from ERA-20C is statistically different from the estimates recovered from climate models and other reanalyses (not shown).

4. Conclusions

Using a subset of CMIP5 models, we evaluate the detectability of externally forced changes in tropical North Atlantic PI for the period 1958–2005 using all-forcing runs as a proxy for observations. This "perfect model scenario" was chosen to eliminate complications from model errors and from observational errors/gaps. Consistent with the findings of Ting et al. (2015) and Sobel et al. (2016), the forced response to greenhouse gas warming and anthropogenic aerosol cooling were found to be nearly collinear and therefore difficult to separate, so we analyzed only the forced runs in which all forcings are combined. An optimization procedure was applied to find the linear combination of large-scale patterns (i.e., Laplacian eigenvectors) in the tropical Atlantic that maximized detectability. Most models (7 of 11) were found to have a detectable signal with only one Laplacian eigenvector. Moreover, including additional Laplacian eigenvectors did not significantly change the projected time series, so we focused the remainder of our analysis on the first Laplacian eigenvector and on the models that had a detectable forced response in this component. Detectability is sensitive to length of the PI record, when the analysis was repeated for the period 1979–2005, the results were consistent, but the number of models detecting a forced response drops to 4 of 11. To allow for comparison,

the time series of the forced response is decomposed into contributions from a linear trend, a quadratic term, and a cubic term. For models with a detectable change in PI, at least one of the features (i.e., trend, quadratic, cubic) characterizing the evolution of the forced response is statistically significant. The trends in these time series do not agree across models: four models have no significant trend, two models have a significant positive trend, and one model has a significant negative trend. The sign of the trend is not even consistent across reanalyses, although the associated confidence intervals are large. A similar level of disagreement is found for the cubic term in the estimated forced response time series. Generally, there is consistency in estimates of PI internal variability among the models and estimates obtained from JRA-55 and NCEP. The magnitude of internal variability found in ERA-20C is dominated by high-frequency variations not found in other reanalysis products or in models. Discrepancies in the trend of PI found in the reanalyses has been attributed to differences in outflow temperature (or tropopause temperature) and surface air-sea disequilibrium (a process largely controlled by SST). For example, Wing et al. (2015) found that 30–36% of the trend in North Atlantic PI in different reanalyses is attributable to differences in tropopause cooling trends. These differences among reanalysis data sets may indicate that hurricane potential intensity is sensitive to the type of variables assimilated or the assimilation scheme itself. The origins of the cross-model differences are currently indeterminate, and a detailed analysis is needed to understand what processes are responsible for the different model responses. The externally forced response in PI captured by the models is weak and in most cases the response cannot be detected even when a single ensemble member is used as a proxy for observations. These results strongly suggest that formal attribution of the forced response of PI in the tropical North Atlantic is not currently feasible, because there is too much model disagreement about the total response and reanalysis products disagree too much in their estimates of low-frequency PI variability. It is possible that as GHG concentrations continue to increase, an unequivocal forced response in North Atlantic PI may emerge in the future.

Acknowledgments
This research was supported primarily by the National Science Foundation (AGS-1622295). Additional support was provided from National Science Foundation (AGS-1338427), National Aeronautics and Space Administration (NNX14AM19G), the National Oceanic and Atmospheric Administration (NA14OAR4310160, NA15OAR43100095, NA16OAR4310079, NA18OAR4310277). The views expressed herein are those of the authors and do not necessarily reflect the views of these agencies. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table S1 in the supporting information) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The various reanalyses are available for download via the Research Data Archive at the National Center of Atmospheric Research from `https://rda.ucar.edu/`. The climate model data are made available through the Coupled Model Intercomparison Project and can be downloaded from https://cmip.llnl.gov/cmip5/data_portal.html. The codes used to estimate the hurricane potential intensity are provided by Kerry Emanuel and can be downloaded here https://emanuel.mit.edu/products.

References

Bengtsson, L. H. K. I., Monika, E., Keenlyside, N., Luis, K., Jing-Jia, L., & Yamada, T. (2007). How may tropical cyclones change in a warmer climate? *Tellus A*, *59*(4), 539–561. https://doi.org/10.1111/j.1600-0870.2007.00251.x

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Chapter 10 Detection and attribution of climate change: From global to regional. In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 867–952). Cambridge: Cambridge University Press.

Bister, M., & Emanuel, K. A. (1998). Dissipative heating and hurricane intensity. *Meteorology and Atmospheric Physics*, *65*(3), 233–240. https://doi.org/10.1007/BF01030791

Camargo, S. J., Ting, M., & Kushnir, Y. (2013). Influence of local and remote SST on North Atlantic tropical cyclone potential intensity. *Climate Dynamics*, *40*(5), 1515–1529. https://doi.org/10.1007/s00382-012-1536-4

Camargo, S. J., & Wing, A. A. (2016). Tropical cyclones in climate models. *WIREs Climate Change*, *7*(2), 211–237. https://doi.org/10.1002/wcc.373

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597.

DelSole, T., & Tippett, M. K. (2007). Predictability: Recent insights from information theory. *Reviews of Geophysics*, *45*, RG4002. https://doi.org/10.1029/2006RG000202

DelSole, T., & Tippett, M. K. (2015). Laplacian eigenfunctions for climate analysis. *Journal of Climate*, *28*(18), 7420–7436. https://doi.org/10.1175/JCLI-D-15-0049.1

Elsner, J. B., Kossin, J. P., & Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature*, *455*, 92–95.

Emanuel, K. A. (1987). The dependence of hurricane intensity on climate. *Nature*, *326*, 483–485.

Emanuel, K. (2000). A statistical analysis of tropical cyclone intensity. *Monthly Weather Review*, *128*(4), 1139–1152. https://doi.org/10.1175/1520-0493(2000)128<1139:ASAOTC>2.0.CO;2

Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, *436*, 686–688.

Emanuel, K., Solomon, S., Folini, D., Davis, S., & Cagnazzo, C. (2013). Influence of tropical tropopause layer cooling on Atlantic hurricane activity. *Journal of Climate*, *26*(7), 2288–2301. https://doi.org/10.1175/JCLI-D-12-00242.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year re-anlysis project. *Bulletin of the American Meteorological Society*, *77*, 437–471.

Klotzbach, P. J., & Landsea, C. W. (2015). Extremely intense hurricanes: Revisiting Webster et al. (2005) after 10 years. *Journal Climate*, *28*(19), 7621–7629. https://doi.org/10.1175/JCLI-D-15-0188.1

Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., et al. (2010). Tropical cyclones and climate change. *Nature Geoscience*, *3*, 157–163.

Knutson, T. R., & Tuleya, R. E. (2004). Impact of CO_2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *Journal Climate*, *17*(18), 3477–3495. https://doi.org/10.1175/1520-0442(2004)017<3477:IOCWOS>2.0.CO;2

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kossin, J. P. (2015). Validating atmospheric reanalysis data using tropical cyclones as thermometers. *Bulletin of the American Meteorological Society*, *96*(7), 1089–1096. https://doi.org/10.1175/BAMS-D-14-00180.1

Kossin, J. P., & Camargo, S. J. (2009). Hurricane track variability and secular potential intensity trends. *Climate Change*, *97*(1), 329–337. https://doi.org/10.1007/s10584-009-9748-2

Kossin, J. P., Olander, T. L., & Knapp, K. R. (2013). Trend analysis with a new global record of tropical cyclone intensity. *Journal of Climate*, *26*(24), 9960–9976. https://doi.org/10.1175/JCLI-D-13-00262.1

Maloney, E. D., Camargo, S. J., Chang, E., Colle, B., Fu, R., Geil, K. L., et al. (2013). North American climate in CMIP5 experiments: Part III: Assessment of twenty-first-century projections. *Journal of Climate*, *27*(6), 2230–2270. https://doi.org/10.1175/JCLI-D-13-00273.1

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, *29*(11), 4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Polvani, L. M., Camargo, S. J., & Garcia, R. R. (2016). The importance of the Montreal Protocol in mitigating the potential intensity of tropical cyclones. *Journal of Climate*, *29*(6), 2275–2289. https://doi.org/10.1175/JCLI-D-15-0232.1

Schneider, T., & Held, I. M. (2001). Discriminants of twentieth-century changes in Earth surface temperatures. *Journal of Climate*, *14*, 249–254.

Smith, A. (2018). 2017 U.S. billion-dollar weather and climate disasters: A historic year in context.

Smith, R. K., Montgomery, M. T., & Vogl, S. (2008). A critique of Emanuel's hurricane model and potential intensity theory. *Quarterly Journal of the Royal Meteorological Society*, *134*(632), 551–561. https://doi.org/10.1002/qj.241

Sobel, A. H., Camargo, S. J., Hall, T. M., Lee, C.-Y., Tippett, M. K., & Wing, A. A. (2016). Human influence on tropical cyclone intensity. *Science*, *353*(6296), 242–246. https://doi.org/10.1126/science.aaf6574

Sobel, A. H., Camargo, S. J., Previdi, M., & Emanuel, K. (2018). Aerosol vs. greenhouse gas influences on tropical cyclone potential intensity. *Bulletin of the American Meteorological Society*, *99*, 1517–1519.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experimental design. *Bulletin of the American Meteorological Society*, *93*, 485–498.

Ting, M., Camargo, S. J., Li, C., & Kushnir, Y. (2015). Natural and forced North Atlantic hurricane potential intensity change in CMIP5 models. *Journal of Climate*, *28*(10), 3926–3942. https://doi.org/10.1175/JCLI-D-14-00520.1

Vecchi, G. A., Fueglistaler, S., Held, I. M., Knutson, T. R., & Zhao, M. (2013). Impacts of atmospheric temperature trends on tropical cyclone activity. *Journal Climate*, *26*(11), 3877–3891. https://doi.org/10.1175/JCLI-D-12-00503.1

Walsh, K. J., McBride, J. L., Klotzbach, P. J., Balachandran, S., Camargo, S. J., Holland, G., et al. (2015). Tropical cyclones and climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 65–89. https://doi.org/10.1002/wcc.371

Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H.-R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, *309*(5742), 1844–1846. https://doi.org/10.1126/science.1116448

Wing, A. A., Emanuel, K., & Solomon, S. (2015). On the factors affecting trends and variability in tropical cyclone potential intensity. *Geophysical Research Letters*, *42*, 8669–8677. https://doi.org/10.1002/2015GL066145

Wing, A. A., Sobel, A. H., & Camargo, S. J. (2007). Relationship between the potential and actual intensities of tropical cyclones on interannual time scales. *Geophysical Research Letters*, *34*, L08810. https://doi.org/10.1029/2006GL028581

Yan, X., Zhang, R., & Knutson, T. R. (2017). The role of Atlantic overturning circulation in the recent decline of Atlantic major hurricane frequency. *Nature Communications*, *8*(1), 1695. https://doi.org/10.1038/s41467-017-01377-8