ENVIRONMENTAL RESEARCH
LETTERS

LETTER • **OPEN ACCESS**

Early and widespread emergence of regional warming is robust to observational and model uncertainty

To cite this article: Jonah Shaw and Nathan Lenssen 2025 *Environ. Res. Lett.* **20** 074066

IOP Publishing Environ. Res. Lett. 20 (2025) 074066 https://doi.org/10.1088/1748-9326/ade458

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
17 December 2024

REVISED
16 April 2025

ACCEPTED FOR PUBLICATION
13 June 2025

PUBLISHED
25 June 2025

LETTER

Early and widespread emergence of regional warming is robust to observational and model uncertainty

Jonah Shaw[1,2,*] and **Nathan Lenssen**[3,4]

[1] Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Boulder, CO 80309, United States of America
[2] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, United States of America
[3] Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO 80401, United States of America
[4] Climate and Global Dynamics Laboratory, NSF National Center for Atmospheric Research, Boulder, CO 80305, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: jonah.shaw@colorado.edu and lenssen@mines.edu

Keywords: climate change, detection and attribution, observational uncertainty

Supplementary material for this article is available online

Abstract

The ability to distinguish observed climate change from naturally arising climate variability is one of the most fundamental findings in climate science. Climate change emergence is a key statistic used to quantify the impact of human activities on climate and inform climate change mitigation strategies. Previous studies determining when climate change has emerged, however, do not account for the combined uncertainty from historical observations and climate models. Here, we show that the emergence of regional warming is early, widespread, and robust to observational and model uncertainty. Warming has emerged over more than 90% of the global land surface as of 2020, and had emerged over more than 50% of the earth's land surface by 2000. Emergence occurs despite observational and model uncertainty creating delays of more than 10 years over 75% of the land surface. These conclusions demonstrate that accounting for multiple sources of uncertainty is necessary for robust detection of climate change emergence.

1. Introduction

The emergence of anthropogenic changes to the climate system has been studied extensively for many climate variables at various spatial and temporal scales. Retrospectively detecting or prospectively estimating the year at which a particular change to the climate system will emerge has been a critical piece of climate science and the communication of climate change, particularly by the IPCC assessment reports (Kirtman *et al* 2013, Doblas-Reyes *et al* 2021, Ranasinghe *et al* 2021). The emergence of changes has been extensively studied for regional surface air temperature (Mahlstein *et al* 2011, Hawkins and Sutton 2012, Kirtman *et al* 2013, Tebaldi and Friedlingstein 2013, Doblas-Reyes *et al* 2021, Ranasinghe *et al* 2021), as well as many other variables such as mean precipitation (Giorgi and Bi 2009), drought (Orlowsky and Seneviratne 2013), extremes (King *et al* 2016), and effects on crop yields (Rojas *et al* 2019). As such, the detection of climate change is one of the most fundamental results of climate change research.

Detecting the time of emergence (ToE) of an anthropogenically forced signal requires providing evidence that the statistics of a climate time series are distinguishable from what could be expected in the absence of anthropogenic forcing. A climate signal has emerged when the observations fall outside of pre-industrial internal variability. We approach emergence as a signal detection problem where the signal is the trend in an observed time series and the noise is the expected variability due to internal variability. Signal-to-noise frameworks are robust and easy to interpret, but are also less appropriate for extreme event detection and can be sensitive to start and end years. We note that emergence can also be thought of as a distribution change (e.g King *et al* 2015) or a change in the likelihood of events (e.g. Stott *et al* 2004). While detection methods have various advantages and yield different results, the underlying

principle is the same: a climate signal emerges when it is distinguishable from internal variability.

Reducing emergence to a signal-to-noise problem means that two key properties of the climate system must be quantified. First, the climate systems' internal variability must be determined. That is, 'Which evolutions of the system could we reasonably expect in the absence of anthropogenic forcing?' Over the past decade, large ensembles of climate models and long pre-industrial control simulations have allowed the characterization of this internal variability (Deser *et al* 2012, 2020, Kay *et al* 2015). Second, the true evolution of the climate system must be quantified for the variable and timescale of interest or 'What evolution of the system actually occurred?' Long, high quality observational records supply this information. Historically, much of the attention in the emergence literature has focused on quantifying model uncertainty in internal variability and the forced response (e.g. Hawkins and Sutton 2012). However, there is substantial uncertainty in observed climate trends, particularly at regional scales (Osso *et al* 2023). There has yet to be a comprehensive study of how both observational and model uncertainty affect ToE.

Various methods have been used to determine when a climate change signal has emerged from internal variability. A common approach is to use climate models to quantify both the internal variability and the climate change signal. Model-based methods generally compute statistics on a running window of fixed length and detect emergence when a window can be distinguished from internal variability (e.g. Hawkins and Sutton 2012, King *et al* 2015). Other studies use a separate set of methods that compute both trends and variability from observed records (Tiao *et al* 1990, Weatherhead *et al* 1998, Leroy *et al* 2008). These observation-based methods detect change in an observed trend of increasing duration rather than from the statistics of a moving window. Model-based methods better characterize internal variability, but are subject to model bias. Observation-based methods, on the other hand, use the true state of the climate system but struggle to appropriately characterize internal variability due to observational uncertainty and relatively short data records.

To isolate the influence of observational and model uncertainty, this work uses a single ToE method. Specifically, we use a hybrid approach similar to Stouffer *et al* (1994) and Shaw and Kay (2023) that combines the advantages of model and observation data sources (see Methods sections 2.2.3 and 2.2.4). We derive internal climate variability from the modeled pre-industrial climate, obtain the climate change signal from observations, and determine detection using linear trends. This trend-based method leverages the long historical record and is well-suited to studying how the uncertainty and availability of observational data influences emergence.

Historical surface temperature anomalies have one of the highest quality climate records over the instrumental era, with near global coverage extending back into the 19th century. Despite the long duration of surface temperature observations, these records also have uncertainties arising from sparse spatial coverage and measurement bias, among others (Chan *et al* 2023, Sippel *et al* 2024). Multiple global temperature products (e.g. GISTEMP, HadCRUT) now provide comprehensive uncertainty assessments that include proper statistical treatment of spatial and temporal uncertainty structure (Huang *et al* 2020, Morice *et al* 2021, Lenssen *et al* 2024). These observational uncertainty ensembles contain multiple, equally likely reconstructions of observed historical surface temperature anomalies and are the best-practice way to quantify observational uncertainty that has statistical properties not well represented by simple covariance structures.

Here, we use uncertainty ensembles to quantify the influence of observational uncertainty on the ToE of regional warming over the global land surface. First, we investigate the spatial and temporal patterns of ToE and delays due to observational uncertainty. We then evaluate individual regions to understand the effects of forced change, internal climate variability, and observational uncertainty on ToE. Finally, we repeat our analysis using additional model and observational data products to create a broad accounting of uncertainty in ToE. To our knowledge, this is the first ToE study to quantify observational uncertainty and combine it with climate model uncertainty. Despite this rigorous and conservative assessment of uncertainty, we find that surface warming had emerged over more than 50% of the land surface by 2000 and has emerged over virtually the entire land surface as of 2020.

2. Data and methods

2.1. Data

2.1.1. Historical surface temperature

We use two historical surface temperature products that estimate global and regional temperature from *in situ* observations: the GISTEMPv4 uncertainty ensemble (Lenssen *et al* 2024) and the HadCRUT5 uncertainty ensemble (Morice *et al* 2021). These products both provide uncertainty ensembles, each containing 200 members that span the known sources of observational uncertainty. Over the land surface, uncertainty arises as a result of correcting inhomogeneities in single weather station as well as the statistical interpolation necessitated by limited spatial and temporal observational coverage (see supplement section 1 for detailed discussion). The two products have different methods and data sources, but broadly agree on regional and global scales (Lenssen *et al* 2024).

We also compare GISTEMP and HadCRUT5 with the Berkeley Earth Land/Ocean Temperature Record (BEST) (Rohde and Hausfather 2020). BEST has a single deterministic member and interpolates more extensively than either GISTEMP or HadCRUT, leading to smoother data with greater spatial and temporal coverage. These three data products do not provide spatially and temporally complete estimates of historical temperature; the treatment of the missing data is discussed below in sections 2.2.1–2.2.3.

Reanalyses also provide valuable records of historical surface temperature, but begin too late in the 20th century to be useful for our analysis. However, we include a comparison of surface temperature trends since 1980 (supplementary figure S1) to show how the ERA5 and MERRA2 reanalysis products compare to the GISTEMP, HadCRUT5, and BEST observational products.

2.1.2. Pre-industrial control runs

We estimate the range of potentially observed trends in a pre-industrial climate for a given region using long pre-industrial control (piControl) simulations from Earth System Models. With dozens of models available from the Coupled Model Intercomparison Project (CMIP), it is tempting to use every model to sample model uncertainty. Applying this 'model democracy', however, ignores the known differences in model performance that can be used to constrain their applications (Hausfather *et al* 2022).

Here, we use four Earth System Models from CMIP5 and CMIP6 that best replicate the observed statistics of regional internal variability over the historical period (Suarez-Gutierrez *et al* 2021, figure 5): CESM1 (Kay *et al* 2015), MPI-ESM-LR (Maher *et al* 2019), CanESM2 (Kirchmeier-Young *et al* 2017), and GFDL-ESM2M (Rodgers *et al* 2015). Such an assessment of climate model internal variability requires a large ensemble over the historical period. The length of the pre-industrial simulations from each model is 500 years (GFDL-ESM2M), 1000 years (CanESM2), 1400 years (CESM1), and 2000 years (MPI-GE). We first present results with CESM1 and apply our methods to all four models in section 5.

2.2. Methods

2.2.1. Annual averages

Calculating annual means from monthly anomalies is the first step in the observed trend analysis. The absence of individual months within a single year can bias an annual mean due to strong seasonality in surface temperatures. To avoid these biases while including as much data as possible, we take a conservative version of the GISTEMP procedure for annual averages (Lenssen *et al* 2019). Annual means are computed when all 4 seasons are properly sampled as defined by 2 of 3 months observed. Annual means are then computed as the average of all 4 seasons so that missing months do not lead to the over-weighting of seasons with more observations.

2.2.2. IPCC region averages

Both observational and model data must be aggregated to the IPCC regions. We use the `xagg` python package move from the raster lat-lon grid to the shapefile IPCC grid and properly account for grid-cell area (Schwarzwald and Geil 2024). The xagg python package was designed to aggregate values on polygons from raster grids, properly accounting for relevant surface areas of both grids. The package calculates the values of the polygonal IPCC regions by a weighted average of overlapping lon/lat raster cells. The weighted mean calculation accounts for the area covered by the raster cell as well as the border cases when raster cells are only partially covered an IPCC region.

Differences in data availability can hinder fair comparisons between observational products, but are also a source of structural uncertainty that wish we to quantify. Consequently, we allow data availability to differ between observational products, but apply conservative methods when computing regional averages to ensure that our results are not biased by incomplete spatial coverage. IPCC regions with data from $<$90% of their area for a timestep are considered unobserved at that timestep. We test the sensitivity of the determined start year and ToE to the regional data availability threshold (section 2.2.3) in supplement section 2 and supplementary figures S2 and S3.

2.2.3. Calculation of observed temperature trends

Trends are calculated using standard linear regression. We compute trends for the first 3 years of a record, then additional years are individually added and trends are recomputed. Unavailable years are treated as missing when trends are computed and not included in the regression. If the final year of a period is unavailable, the trend for that year is masked when computing ToE so that emergence does not occur in the absence of observations. That is, we prevent a trend emerging because the envelope of internal variability narrows enough that the trend from the preceding year emerges.

In addition to using a conservative data availability threshold when computing regional averages (section 2.2.2), we also apply a conservative approach for determining the start year of the trend analysis. Specifically, we start the trend analysis at the first year in the 20th century with consistent observations thereafter. For each spatial region (IPCC and lat-lon grid), we identify this year as the earliest year when: (1) Data is available for that year and (2) the following 5, 10, 20, 40, and 80 year periods all have at least 66% data availability. The results presented are robust to changes in this method.

2.2.4. Internal climate variability

The possible range of trends in a pre-industrial climate is computed by sampling each piControl simulation for time series of lengths from 3 to 120 years. All possible series of a given length are sampled and the linear trend is calculated. Confidence intervals are then calculated empirically from the sampled trends for each of the four Earth System Models used.

While pre-industrial control simulations are commonly used to estimate internal climate variability for detection studies (e.g. Stouffer *et al* 1994, Hegerl *et al* 1996), they do not include non-anthropogenic forcings such as volcanic eruptions and variations in the solar constant. Such forcings may increase the range of pre-industrial trends, but these effects should be small on the multi-decadal timescales at which we detect trend emergence (e.g. Lockwood 2012). By computing trends from the beginning of consistent observations (see section 2.2.3), we ensure that the influence of non-anthropogenic forcings on our results is small.

2.2.5. Time of emergence

We define a median and strict ToE to assess the influence of uncertainty on ToE. For both metrics, we compute ToE by comparing observed trends with a 95% confidence interval on unforced trends. The median ToE is defined as the year which the median observed trend exceeds the upper 95% confidence interval of pre-industrial trends and stays above it for the remainder of the record. The 'median ToE' is approximately equivalent to calculating the ToE with a single deterministic temperature record. The strict ToE is calculated by first determining the ToE for each observational uncertainty ensemble member separately. As in the median definition, ToE for each member is defined as the year when a trend exceeds the ICV interval and does not re-enter. Then, we define the strict ToE as the year when when 95% of the observational ensemble members have emerged. By construction, the 'strict ToE' will be the same or a later year than the 'median ToE' in line with the intuition that observational uncertainty impairs our ability to detect change in the climate system.

2.2.6. Total uncertainty accounting

Using two observational models, four models, and 200 ensemble members results in 1600 estimates of ToE for each region. This ToE distribution samples observational uncertainty, model uncertainty, and observational product uncertainty. We apply our definitions of median and mean ToE to this entire distribution to produce a conservative estimate of ToE at both IPCC Region and lat-lon grid scale.

2.2.7. Time of emergence delays

We compute delay in ToE due to observational uncertainty alone (Observational Delay) by taking the difference between the median and strict ToE for a single observational uncertainty data product (GISTEMP or HadCRUT5). We compute delay due to combined observational uncertainty, model uncertainty, and observational product uncertainty (Total Delay) analogously to the observational delay while using the ToE distribution described in section 2.2.6.

2.2.8. Spatial scale

The spatial scale at which emergence is studied largely determines the relevant audience of the results. Emergent trends over large regions defined by the IPCC provide useful global summaries (Iturbide *et al* 2020). Trend detection at finer scales where climate impacts and adaptation efforts occur, however, is more relevant to local climate policy and communication. At these finer scales, however, internal variability is larger and emergence generally later (Shaw and Kay 2023). We address the question of scale by computing ToE at both the IPCC region scale on the native 5 degree lat-lon grid of available uncertainty ensembles.

3. Effect of observational uncertainty on trend emergence

We first determine how observational uncertainty affects ToE for the land surface IPCC regions as these regions are the primary form that emergent trends are communicated to the broader climate change community (Iturbide *et al* 2020). GISTEMP and CESM1 (Kay *et al* 2015) are used to compute observed trends and internal climate variability, respectively. In agreement with previous findings, temperature trends have emerged above internal variability for nearly all land surface IPCC regions (figures 1(a) and (b)). The only IPCC regions that have not emerged are in the Antarctica, due to insufficient data availability (see Methods 2.2.2). Observational uncertainty delays ToE by more than five years for almost all IPCC regions (figure 1(c)), and 20% of the Earth's surface has delays greater than 15 years.

Regions with large delays are predominantly in the tropics and the southern hemisphere where observational uncertainty is generally greater due to fewer observations (Menne *et al* 2018, Lenssen *et al* 2024). Smaller forced trends in the tropics also delay emergence (Hawkins and Sutton 2012). Supplementary figures S4(a)–(c) compares ToE for the GISTEMP, HadCRUT5, and BEST observational products.

We next investigate the emergence of trends on a regular lat-lon grid instead of pre-defined regions (figures 1(d)–(f)). Similar to the IPCC regions, trends have emerged over nearly the entire non-Antarctic surface even after accounting for observational uncertainty (figures 1(d) and (e)). The greatest observational delays are consistent with the IPCC regions, but the gridded analysis better resolves regions where trends emerge early and others where observational uncertainty causes large delays.

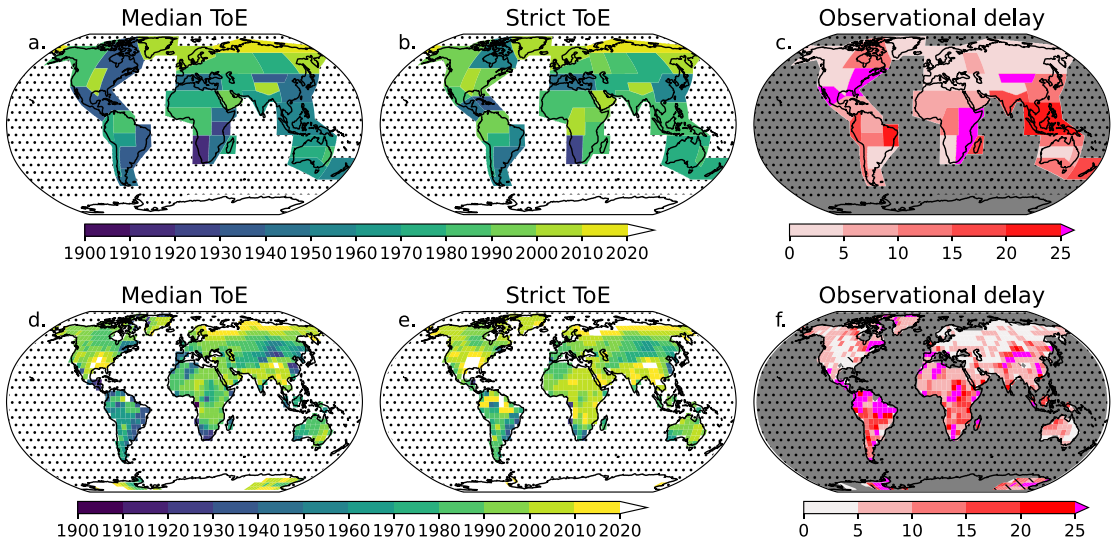

Figure 1. Time of Emergence (ToE) for GISTEMP and CESM1. (a) The ToE at IPCC region spatial scale without taking observational uncertainty into consideration (Median ToE). (b) The updated ToE estimate, taking into account observational uncertainty (Strict ToE). (c) The observational delay, defined as the difference in the two ToE years. (d)–(f) as (a)–(c), but for a 5 degree lat-lon grid. Dotted hatching indicates ocean surfaces. Slashed hatching indicates regions where strict ToE has not occurred and delay is a lower bound.

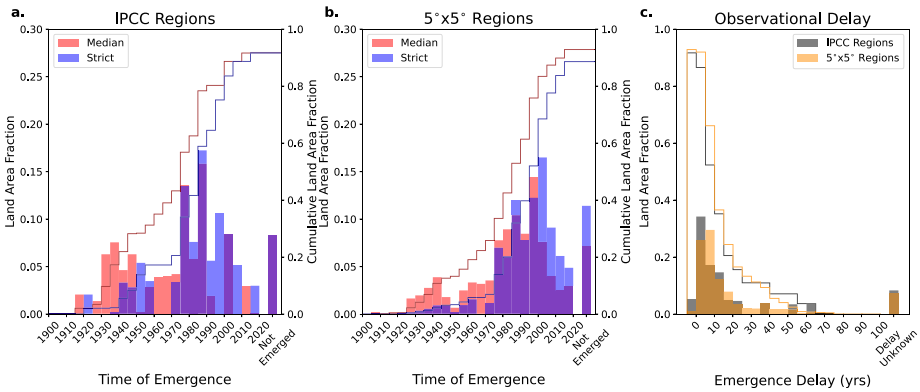

Figure 2. Histograms (filled) and cumulative distribution functions (unfilled) of the Time of Emergence (ToE) year under the median and strict ToE definitions for (a) IPCC regions and (b) a 5 degree lat-lon grid. (c) The delay due to including observational uncertainty. Unfilled values in panel (c) are 1 minus the cumulative distribution function, indicating the land surface fraction with delays greater than or equal to the *x*-axis delay value. ToE is computed using CESM1 and GISTEMP.

Well-sampled regions of low internal variability on the Brazil coast of South America and in Southern Africa emerge by 1950 or earlier even after accounting for observational uncertainty (figure 1(e)). On the other hand, nearby regions in the Amazon and Central Africa emerge decades later after delays of more 25 years (figure 1(f)). In general, the regions with long delays are concentrated in the Global South where observational uncertainty is large. Supplementary figures S4(d)–(f) compare ToE for the GISTEMP, HadCRUT5, and BEST observational products.

While the IPCC regions and lat-lon grid may agree qualitatively, the ToE occurs later for the lat-lon regions when compared to the larger IPCC regions (figures 2(a) and (b)). This result is consistent for both the median ToE and strict ToE as the ratio between the forced signal and internal variability is expected to decrease as the area of a region decreases. Regardless of the spatial scale used, however, observational uncertainty delays the ToE (figures 1 and 2).

Delays in ToE are also generally greater on the lat-lon grid than the IPCC Region scale (figure 2(c)). 68% (58%) of lat-lon (IPCC) regions are delayed by more than 5 years. Despite these large delays, the detection of regional warming is widespread. As of 2020, 91% of the Earth's land surface area has an emergent signal of climate change when accounting for observational uncertainty at the IPCC regional spatial scale. On the finer lat-lon grid, 81% of the land surface has emerged when accounting for observational uncertainty. Supplementary figure S5 shows Median ToE histograms at the IPCC region and 5 degree scale for the HadCRUT5 and BEST observational products.

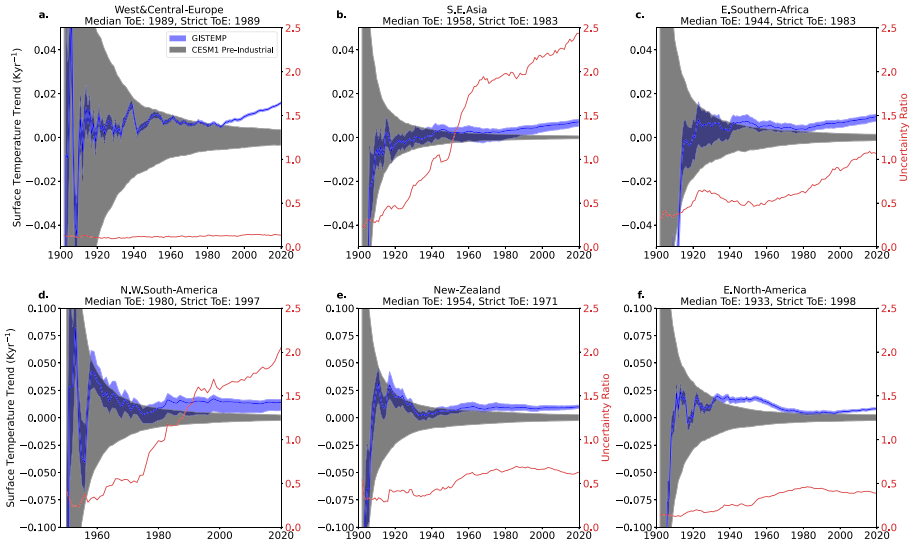

Figure 3. Time of Emergence plots for six IPCC regions. Observed regional trends (dark blue), their corresponding observational uncertainty (light blue), and the 95% likely pre-industrial trends (grey). The ratio between the 95% likely ranges of the observational uncertainty and internal variability (red) should be approximately constant if the observational uncertainty and internal variability each have constant uncertainty in time. In panel (d), trends begin in 1948 due to data availability (see Methods 2.2.2). ToE is computed using CESM1 and GISTEMP.

4. Variations in delay of emergence

We next investigate 6 IPCC regions and demonstrate the interplay between observational uncertainty, trends, and internal climate variability (figure 3). Regions are selected from each continent except Antarctica. The West and Central Europe region (figure 3(a)) has low observational uncertainty and a fairly monotonic warming signal, and is the only region of the six that has no delay in ToE. The other five regions have delays ranging from 17 to 65 years (figures 3(b)–(f)). These five regions of large delay provide illustrative examples of how observational uncertainty and internal variability can alter ToE in non-intuitive ways. First, all five regions show a moderation or decrease in the observed trend shortly after the median trend emerges. The median trend remains detectable, but observational ensemble members with less warming do not emerge until much later. The result leads to two conclusions: ToE would likely be much earlier if surface temperature were known more confidently, and internal variability superimposed on the underlying forced response created conditions for earlier emergence than if trends followed the monotonic forced response. The inclusion of observational uncertainty also prevents early false alarms of emergence due to multi-decadal variability as is seen in Eastern North America (figure 3(f)).

While the five regions of large delay show somewhat consistent behavior in trends following emergence, they differ markedly in how observational uncertainty evolves in time. Specifically, S.E. Asia and N.W. South America show a large increase in observational uncertainty relative to internal variability (figures 3(b) and (d), red lines). These two regions are in sparsely observed areas where observational uncertainty in monthly temperature anomalies has increased in recent decades (Lenssen *et al* 2024). In the well sampled West and Central Europe, observational uncertainty remains flat relative to internal variability as theorized by the underlying time series statistics assuming autoregressive time series (Leroy *et al* 2008). The examples provided here give insight into the interplay between observational uncertainty and internal variability.

These examples highlight that there are three properties of a region that control the ToE: (1) The magnitude and variability of the observed trend, (2) the magnitude and temporal structure of the observational uncertainty and (3) the magnitude of the pre-industrial internal variability. Here, we have showed that large delays primarily occur due to some combination of factors (1) and (2); the largest delays are seen in regions with non-monotonic trends and/or persistent observational uncertainty. These two factors have a much larger influence on the delay of ToE than does the magnitude of pre-industrial variability as shown by the presence of large delays in regions with both low and high trend internal variability (McKinnon and Deser 2018). We find the examples provided here give insight into the complex interplay between the various uncertainties and guidance for possible considerations in the development of ToE detection methods.

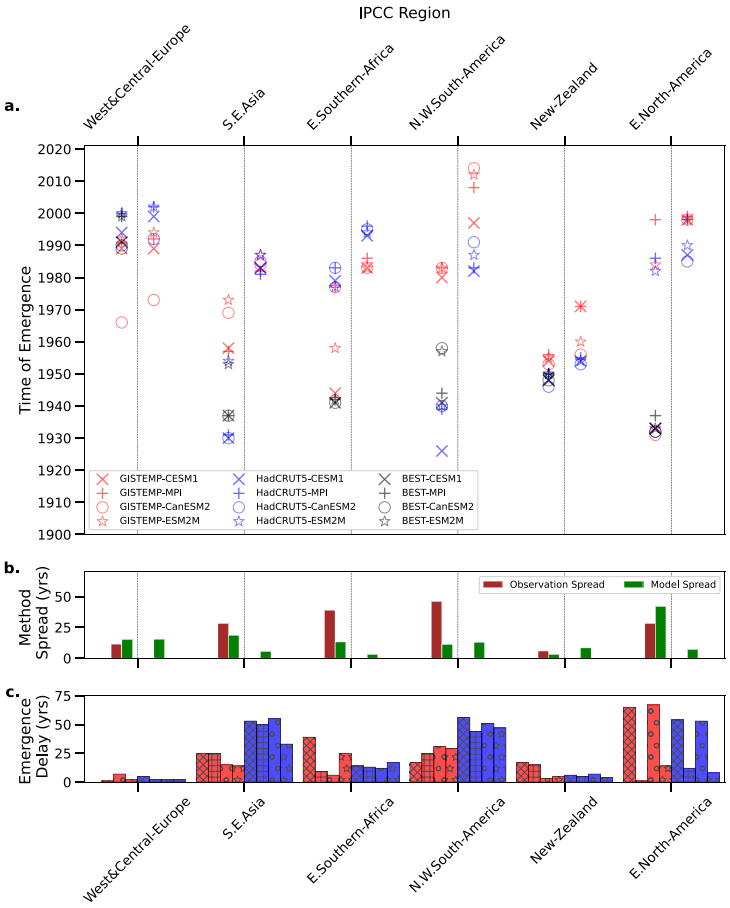

Figure 4. Spread in median and strict ToE for six selected IPCC regions. (a) For each region, the median (strict) ToE for each model-observation data pairing is reported on the left (right) of the vertical dashed line. Symbol pattern and color indicates the model and observational data product, respectively. (b) Spread in median (left) and strict (right) ToE due to choice of model and observational product. Observational spread is the average change in ToE when the observational product is changed with model product held constant. Model spread is the average difference between maximum and minimum ToE across model products when the observational product is held constant. (c) ToE delay for each of the model-observation pairing with symbols and colors as in panel (a). ToE computed from BEST is compared with median ToE from the GISTEMP and HadCRUT5 ensembles.

5. How much does the choice of data product matter?

Any estimate of ToE is subject to epistemic uncertainty from unquantified error in model and observational data products. To estimate the sensitivity of the presented results to the choice of temperature product, we repeat our analysis with the Met Office Hadley Centre/ Climatic Research Unit global surface temperature dataset version 5 (HadCRUT5) (Morice *et al* 2021). To sample the epistemic uncertainty in model estimates of internal climate variability, we repeat our analysis with an additional three Earth System Models (MPI-ESM, CanESM2, GFDL-ESM2M), selecting climate models that best represent internal variability in the study of Suarez-Gutierrez *et al* (2021). Thus, we repeat our analysis eight times by crossing the two temperature products with four estimates of internal variability.

Figure 4 shows the median and strict ToE from each model-observation pairing for the 6 IPCC regions discussed in section 4, allowing us to examine the influence of different data products. There are clear differences in model internal variability in E. North America, which give rise to large delays when CESM1 or CanESM2 are used. Conversely, ToE is highly sensitive to the choice of observational product in N.W. South America where GISTEMP shows greater observational uncertainty than HadCRUT5 (Lenssen *et al* 2024). Overall, figure 4 paints a complex picture of epsitemic uncertainty where there is no persistent pattern among model-observation pairings. This result gives us confidence that this method of quantifying epistemic uncertainty is doing a reasonable job of summarizing the uncertainty in ToE estimates due to methodological choices. ToE results for all IPCC regions and model-observation pairings are included in the supplement (figures S6–S8).

Maps of epistemic uncertainty in strict ToE on the lat-lon grid (figure 5) also show a complex pattern. In general, spread in strict ToE is mostly due to the choice of observational data product (figures 5(a) and (b)), which aligns well with regions that have large uncertainty in historical surface temperature. In South America and Central Asia, for example, strict ToE varies by over 40 years. Still, recovering

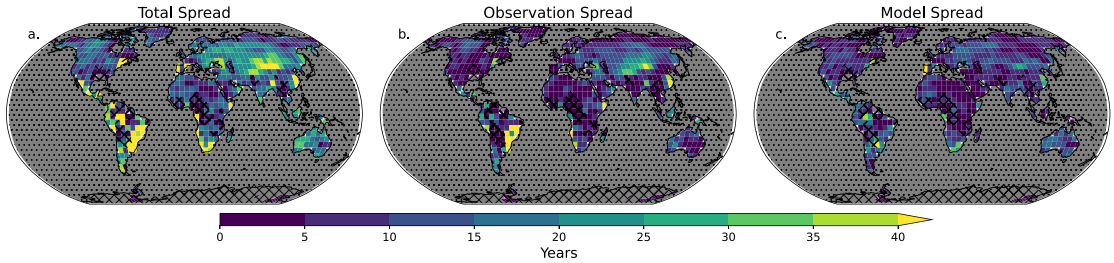

Figure 5. (a) Total spread in ToE on the 5 degree lat-lon grid. (b) Spread due to choice of observational data product. (c) Spread due to choice of model product. Observational spread is the average change in ToE when the observational product is changed with model product held constant. Model spread is the average difference between maximum and minimum ToE across model products when the observational product is held constant. Dotted hatching indicates ocean surfaces. Slashed hatching indicates land surfaces where spread is calculated using fewer than four model-observation pairings due to missing ToE values.

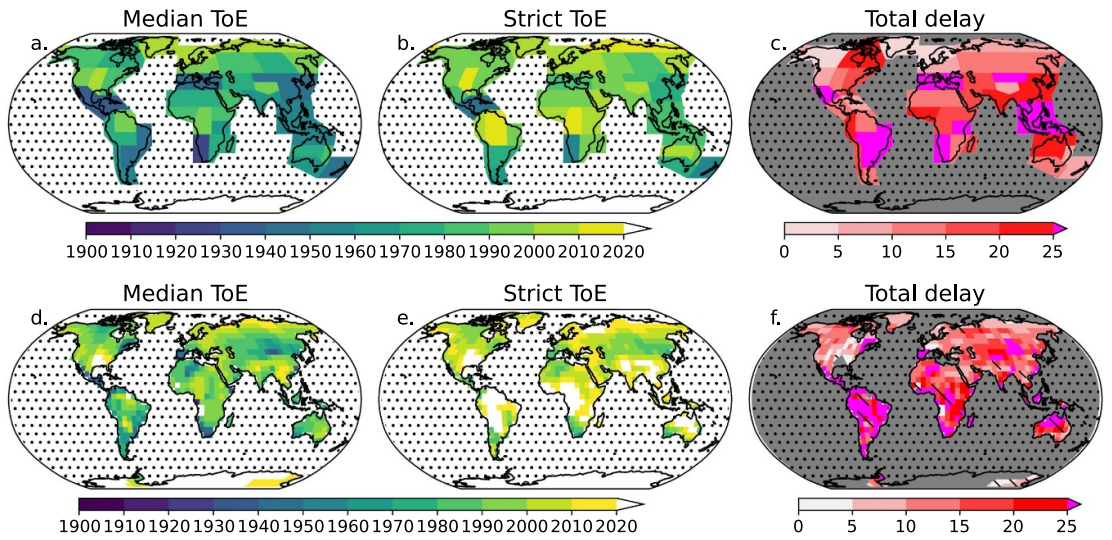

Figure 6. Time of emergence (ToE) when accounting for observational uncertainty, model uncertainty, and observational product uncertainty. (a) Median ToE, (b) strict ToE, and (c) the total delay, defined as the difference in the two ToE years. Panels (d)–(f) are as in (a)–(c), but for a 5 degree lat-lon grid. Dotted hatching indicates ocean surfaces. Slashed hatching indicates regions where strict ToE has not occurred and delay is a lower bound.

the total spread in strict ToE requires sampling across both observational and model products, demonstrating that each data source is an important source of epistemic uncertainty.

Finally, we combine all sources of uncertainty examined here to produce a conservative estimate of ToE (figure 6). Using two observational models, four models, and 200 ensemble members, we produce 1600 estimates of ToE for each region. Requiring that 95% of these estimates emergence guarantees a robust and conservative assessment of surface warming detection. At IPCC region scale, the entire non-Antarctic land surface has emerged, and more than 50% of the total land surface had emerged by 2000 (figure S9(a)). On the lat-lon grid, 74% of the total land surface has emerged and more than 50% had emerged by 2010 (figure S9(b)). This widespread and early emergence occurs despite large delays due to uncertainty. At IPCC (lat-lon grid) scale, 75% (68%) of the land surface is delayed by more than 10 years, and 35% (30%) is delayed by more than 20 years.

6. Conclusions

In this study, we have expanded the calculation of ToE to include the uncertainty in historical surface temperature, as well as epistemic uncertainty arising from the choice of data products used. This treatment of relevant uncertainties in calculating ToE provides guidance for incorporating uncertainty in future studies, including IPCC reports. In addition, combining multiple uncertainty sources to investigate regional ToE has shown that:

(1) Forced surface temperature trends have emerged for nearly the entire land surface globe. More than 50% of the land surface had emerged by 2000 and 90% had emerged as of 2020 (IPCC scale).
(2) This robust detection occurred despite large delays due to observational and model uncertainty; more than 35% of the land surface is

delayed by more than 20 years (IPCC scale). Median ToE occurs decades earlier, suggesting that reducing observational and climate model uncertainty would lead to widespread regional climate change emergence during the 20th century.

(3) Observational uncertainty is the largest source of delays. The concentration of large delays in the Global South provides a striking example of how historical observational practices become embedded in critical IPCC science (e.g. Lieber *et al* 2022).

(4) Non-monotonic warming trends and decreases in the quality of observational records cause the greatest delays in ToE.

All together, these results provide the most statistically complete assessment of regional surface warming detection to date. Despite the inclusion of additional uncertainty, the detection of regional warming is early and widespread.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.15686231.

Acknowledgment

The authors thank J E Kay and E Hawkins as well as three anonymous reviewers for comments on the manuscript. J. K. S. was supported by the NASA PREFIRE mission Award 849K995 and NASA FINESST Grant 80NSSC22K1. N. L. is partially funded through NCAR which is sponsored by the National Science Foundation under Cooperative Agreement 1852977. Computing and data storage resources, including the Cheyenne (doi:10.5065/D6RX99HX) and Derecho (doi:10.5065/qx9a-pg09) supercomputers, were provided by the National Science Foundation's National Center for Atmospheric Research (NSF NCAR)'s Computational and Information Systems Laboratory.

ORCID iDs

Jonah Shaw https://orcid.org/0000-0003-3479-4032
Nathan Lenssen https://orcid.org/0000-0002-6601-4187

References

Chan D, Gebbie G and Huybers P 2023 Global and regional discrepancies between early-twentieth-century coastal air and sea surface temperature detected by a coupled energy-balance analysis *J. Clim.* **36** 2205–20

Deser C *et al* 2020 Insights from earth system model initial-condition large ensembles and future prospects *Nat. Clim. Change* **10** 277–86

Deser C, Phillips A, Bourdette V and Teng H 2012 Uncertainty in climate change projections: the role of internal variability *Clim. Dyn.* **38** 527–46

Doblas-Reyes F J *et al* 2021 Linking global to regional climate change *Climate Change 2021: The Physical Science Basis Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press) (https://doi.org/10.1017/9781009157896.012)

Giorgi F and Bi X 2009 Time of emergence (TOE) of GHG-forced precipitation change hot-spots *Geophys. Res. Lett.* **36** L06709

Hausfather Z, Marvel K, Schmidt G A, Nielsen-Gammon J W and Zelinka M 2022 Climate simulations: recognize the 'hot model' problem *Nature* **605** 26–29

Hawkins E and Sutton R 2012 Time of emergence of climate signals *Geophys. Res. Lett.* **39** L01702

Hegerl G C, von Storch H, Hasselmann K, Santer B D, Cubasch U and Jones P D 1996 Detecting greenhouse-gas-induced climate change with an optimal fingerprint method *J. Clim.* **9** 2281–306

Huang B *et al* 2020 Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5 *J. Clim.* **33** 1351–79

Iturbide M *et al* 2020 An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets *Earth Syst. Sci. Data* **12** 2959–70

Kay J E *et al* 2015 The Community Earth System Model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability *Bull. Am. Meteorol. Soc.* **96** 1333–49

King A D, Black M T, Min S K, Fischer E M, Mitchell D M, Harrington L J and Perkins-Kirkpatrick S E 2016 Emergence of heat extremes attributable to anthropogenic influences *Geophys. Res. Lett.* **43** 3438–43

King A D, Donat M G, Fischer E M, Hawkins E, Alexander L V, Karoly D J, Dittus A J, Lewis S C and Perkins S E 2015 The timing of anthropogenic emergence in simulated climate extremes *Environ. Res. Lett.* **10** 094015

Kirchmeier-Young M C, Zwiers F W and Gillett N P 2017 Attribution of extreme events in Arctic sea ice extent *J. Clim.* **30** 553–71

Kirtman B *et al* 2013 Near-term climate change: projections and predictability *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press) (https://doi.org/10.1017/CBO9781107415324.023)

Lenssen N, Schmidt G A, Hansen J E, Menne M J, Persin A, Ruedy R and Zyss D 2019 Improvements in the GISTEMP uncertainty model *J. Geophys. Res.: Atmos.* **124** 6307–26

Lenssen N, Schmidt G A, Hendrickson M, Jacobs P, Menne M J and Ruedy R 2024 A NASA GISTEMPv4 observational uncertainty ensemble *J. Geophys. Res.: Atmos.* **129** e2023JD040179

Leroy S S, Anderson J G and Ohring G 2008 Climate signal detection times and constraints on climate benchmark accuracy requirements *J. Clim.* **21** 841–6

Lieber R, King A, Brown J, Ashcroft L, Freund M and McMichael C 2022 ENSO teleconnections more uncertain in regions of lower socioeconomic development *Geophys. Res. Lett.* **49** e2022GL100553

Lockwood M 2012 Solar influence on global and regional climates *Surv. Geophys.* **33** 503–34

Maher N *et al* 2019 The Max Planck Institute grand ensemble: enabling the exploration of climate system variability *J. Adv. Model. Earth Syst.* **11** 2050–69

Mahlstein I, Knutti R, Solomon S and Portmann R W 2011 Early onset of significant local warming in low latitude countries *Environ. Res. Lett.* **6** 034009

McKinnon K A and Deser C 2018 Internal variability and regional climate trends in an observational large ensemble *J. Clim.* **31** 6783–802

Menne M J, Williams C N, Gleason B E, Rennie J J and Lawrimore J H 2018 The global historical climatology network monthly temperature dataset, version 4 *J. Clim.* **31** 9835–54

Morice C P, Kennedy J J, Rayner N A, Winn J, Hogan E, Killick R, Dunn R, Osborn T, Jones P and Simpson I 2021 An updated assessment of near-surface temperature change from 1850: the HadCRUT5 data set *J. Geophys. Res.: Atmos.* **126** e2019JD032361

Orlowsky B and Seneviratne S I 2013 Elusive drought: uncertainty in observed trends and short-and long-term CMIP5 projections *Hydrol. Earth Syst. Sci.* **17** 1765–81

Osso A, Craig P and Allan R P 2023 An assessment of CMIP6 climate signals and biases in temperature, precipitation and soil moisture over Europe *Int. J. Climatol.* **43** 5698–719

Ranasinghe R *et al* 2021 Climate change information for regional impact and for risk assessment *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press) (https://doi.org/10.1017/9781009157896.014)

Rodgers K B, Lin J and Frölicher T L 2015 Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model *Biogeosciences* **12** 3301–20

Rohde R A and Hausfather Z 2020 The Berkeley Earth land/ocean temperature record *Earth Syst. Sci. Data Discuss.* **2020** 1–16

Rojas M, Lambert F, Ramirez-Villegas J and Challinor A J 2019 Emergence of robust precipitation changes across crop production areas in the 21st century *Proc. Natl Acad. Sci.* **116** 6673–8

Schwarzwald K and Geil K 2024 xagg: a Python package to aggregate gridded data onto polygons (available at: https://github.com/ks905383/xagg)

Shaw J K and Kay J E 2023 Processes controlling the seasonally varying emergence of forced Arctic longwave radiation changes *J. Clim.* **36** 7337–54

Sippel S *et al* 2024 Early-twentieth-century cold bias in ocean surface temperature observations *Nature* **635** 618–24

Stott P A, Stone D A and Allen M R 2004 Human contribution to the European heatwave of 2003 *Nature* **432** 610–4

Stouffer R J, Manabe S and Vinnikov K Y 1994 Model assessment of the role of natural variability in recent global warming *Nature* **367** 634–6

Suarez-Gutierrez L, Milinski S and Maher N 2021 Exploiting large ensembles for a better yet simpler climate model evaluation *Clim. Dyn.* **57** 2557–80

Tebaldi C and Friedlingstein P 2013 Delayed detection of climate mitigation benefits due to climate inertia and variability *Proc. Natl Acad. Sci.* **110** 17229–34

Tiao G C, Reinsel G C, Xu D, Pedrick J H, Zhu X, Miller A J, DeLuisi J J, Mateer C L and Wuebbles D J 1990 Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation *J. Geophys. Res.: Atmos.* **95** 20507–17

Weatherhead E C *et al* 1998 Factors affecting the detection of trends: statistical considerations and applications to environmental data *J. Geophys. Res.: Atmos.* **103** 17149–61