Observational limitations to the emergence of climate signals

Louis Rivoire[1,2], Marianna Linz[1,3], Jingyuan Li[4]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA
[2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA
[4]Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA

Key Points:

- The degree of confidence placed in observed climate trends is misrepresented when overlooking observational limitations;
- We provide a nonparametric method to account for such limitations;
- The method can also inform the design of future observing platforms.

Corresponding author: Louis Rivoire, `lrivoire@mit.edu`

Abstract

Using model projections to study the emergence of observable climate signals presumes omniscient knowledge about the climate system. In reality, observational knowledge suffers from data quality and availability issues. Overlooking such deficiencies leads to misrepresentations of the time of emergence (ToE). We introduce a new definition of ToE that accounts for observational limitations (e.g., data gaps, gridding, changes in instrumentation, retrieval algorithms, etc), and show the potential for significant corrections to achieve the same statistical confidence as would be afforded by omniscient knowledge. We also show how our method can inform future observational needs and observing systems design.

Plain Language Summary

Long-term planning for climate change adaptation requires accurate forecasts of climate impacts. Such forecasts are produced using computer models, which provide omniscient knowledge of the climate states they simulate. However, real-world knowledge is based on incomplete and sometimes flawed observational data. Ignoring these flaws yields a distorted view of the timing of observable climate impacts. We propose a method to address this issue by accounting for observational limitations such as data gaps, changes in measuring equipment, data post-processing, etc. We show how to use the method to plan future data collection.

1 Introduction

Despite global ambitions for climate action, adaptation gaps persist. Systemic barriers such as limited climate literacy and data availability (H. Lee et al., 2023) stand in the way of progress, along with the growing challenge of maladaptation for vulnerable groups (Schipper, 2020; Pörtner et al., 2023) which is aggravated by social inequalities (Islam & Winkel, 2017). Addressing these issues will require comprehensive and effective policy packages for long-term adaptation (Biesbroek et al., 2013), which rely on quantitative knowledge of climate trends and risk (Pörtner et al., 2023; National Academies of Sciences & Medicine, 2018).

Knowledge about future climate trends is subject to the limitations of climate models. In the context of trend detection, misrepresentation of decadal to multidecadal internal variability by models is problematic (Collins et al., 2002; Danabasoglu, 2008; Bothe et al., 2013; Kim et al., 2018; Tao et al., 2023) especially on regional scales (Laepple et al., 2023). The scientific community has addressed uncertainties due to internal variability with the use of large ensemble modeling (Zelle et al., 2005; Drijfhout et al., 2008; Branstator & Selten, 2009; Rodgers et al., 2021) and other downscaling or bias-correction techniques (Wu et al., 2022). Inter-model spread has also been addressed using emergent constraints (Williamson et al., 2021; Qasmi & Ribes, 2022).

However, disagreements between models and observations persist (e.g., Abalos et al., 2021). Disagreements can arise for a variety of reasons including the mere presence of internal variability (Jain et al., 2023; Tebaldi & Knutti, 2007; Mitchell et al., 2013; McKinnon & Simpson, 2022). As a result, direct comparison of models with observations is inappropriate (Collins et al., 2013; Schmidt et al., 2023). In response, recent methods integrate both model-based and observational knowledge to better account for internal variability. For instance, McKinnon and Deser (2021) quantify uncertainties related to sampling of internal variability (see also Shepherd, 2021; Gessner et al., 2021; Barnes et al., 2019).

When observed and simulated trends are at odds, questions arise: are the models wrong? Is the observational record long enough, and of high enough quality? If not, how much, and what kind of additional data should be collected? This study addresses the latter two questions using the concept of time of emergence (ToE): the time after which a trend becomes distinguishable from background variability. The concept of ToE is useful to:

1. Incorporate uncertainties due to internal variability in the assessment of climate models;
2. Communicate climate change (e.g. by determining when the effects of climate change will likely manifest to convey the urgency of taking action to stakeholders, the general public, and policymakers);
3. Develop mitigation and conservation strategies (e.g. by providing a timeline for the allocation of resources for research, infrastructure upgrades, disaster preparedness, etc);

Methods already exist to quantify ToE (see Section 2), but they rely on climate model data and exclude observational uncertainties. In this study, we introduce (Section 2) and validate (Section 3) a method to quantify the additional length of record needed to account for observational limitations in the emergence of observed climate signals. The method is useful to analyze historical records and to design future observing systems (Section 4).

2 Methods

The detectability of climate change has long been quantified using the concept of signal-to-noise ratio (Madden & Ramanathan, 1980; Wigley & Jones, 1981; Barnett & Schlesinger, 1987; Santer et al., 1995; Hegerl et al., 1996, 1997; Bindoff et al., 2014; Wills et al., 2020), where the signal is a measure of a trend and the noise one of internal climate variability. Drawing from this concept, ToE has often been defined as the length of record beyond which the signal-to-noise ratio exceeds a predetermined threshold (Christensen et al., 2007; Giorgi & Bi, 2009; Diffenbaugh & Scherer, 2011; Hawkins & Sutton, 2012; Deser et al., 2012; Maraun, 2013; Sui et al., 2014; Lyu et al., 2014; Zappa et al., 2015; D. Lee et al., 2016; Nguyen et al., 2018). Emergence has also been defined using other statistical tests for the difference between a reference state and a perturbed state (Mahlstein et al., 2011), with comparable results.

While these methods provide useful information about climate signals, they differ in their definitions of the signal, noise, and threshold (see a discussion in Li et al., 2017) and suffer from key limitations. For instance, the choice of signal-to-noise threshold is arbitrary and does not provide a standardized definition for statistical confidence. Additionally, the concept of signal-to-noise ratio does not account for autocorrelation in climate time series, leading to underestimated ToE and misrepresented spatial patterns of emergence (Li et al., 2017).

The method developed by Li et al. (2017) (L17 hereafter) addresses these shortcomings by defining emergence as the time when an analytical confidence interval (defined by Thompson et al., 2015) about a cumulative trend excludes zero. However, to our knowledge, no method explicitly accounts for observational limitations. We propose a new definition of ToE to address this shortfall.

2.1 New Definition of Emergence

Applying ordinary least squares regression to a climate time series generates a linear trend estimate, denoted b. To assess whether b constitutes a significant departure from internal variability, we compare it to the distribution of trends that arise purely as a result of internal variability over the same time period. If b is close to the first moment of this reference distribution, then it aligns with typical fluctuations seen in the climate system without external forcings. Conversely, if b falls well into the tails of the reference distribution, then it is unusually large compared to natural variations. In such a scenario, it is reasonable to hypothesize that b may have arisen due to external forcings.

The reference distribution is derived from a control simulation that excludes external forcings: a resampling time window of length y (in time steps) is chosen and linear trends for all possible y-step periods (with overlap) are calculated using ordinary least squares regression. For a control run with monthly outputs, choosing y=120 yields the distribution of 10-year unforced trends. Figure 1b shows reference distributions derived from a synthetic control run (panel a) for two resampling windows. The control run is produced by random number generation based on Pearson distributions (Pearson, 1894; Johnson et al., 1995). White noise is used in this example, but the method works for any power spectrum of noise. As one may expect, the greater the y, the narrower the trend distribution. We define ToE as the resampling window length y needed to obtain a reference distribution such that trends larger than b are statistically unlikely to occur, with a chosen degree of statistical confidence. This definition handles time series with autocorrelation and extends the capabilities of L17 by handling time series with non-Gaussian residuals about the trend. Note that internal variability is assumed to be constant over time, which may be inappropriate for some climate variables (Rodgers et al., 2021). The formalism and procedure are laid out in the next section.

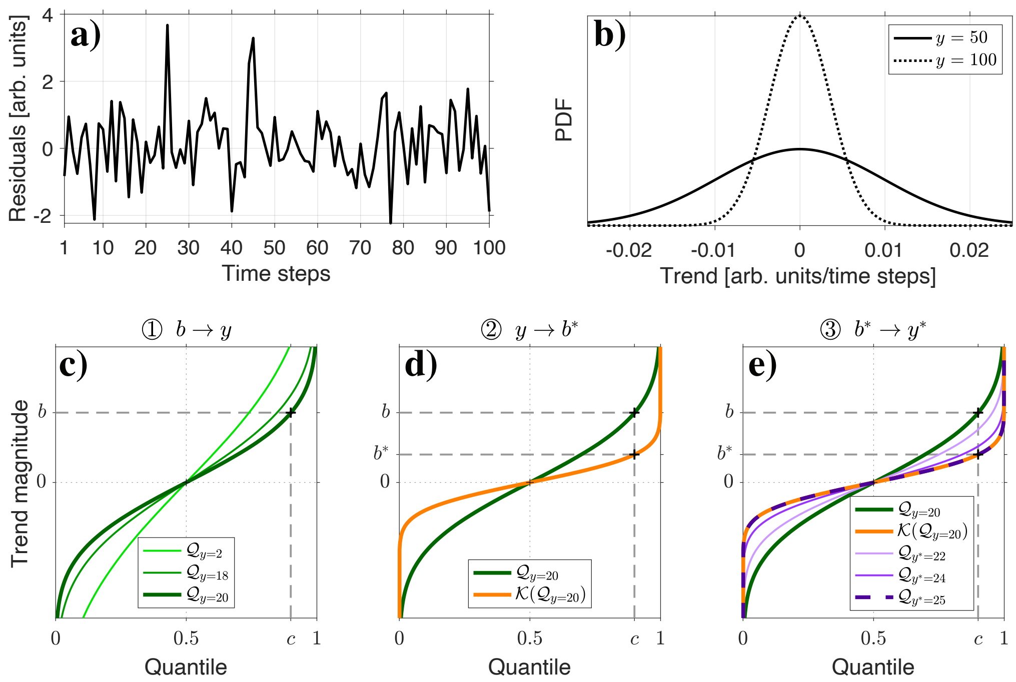

Figure 1. a) Sample time series from a synthetic control simulation and b) corresponding probability density function estimates of linear trends for varying resampling window lengths y (in time steps). c) Illustration of the resampling method and d-e) adjustment of the ToE for observational limitations when $b^* < b$. Light green and light purple curves show the iterative process by which equations 1 and 2 are fulfilled. Numerical values are for illustration purposes only.

2.2 Time of Emergence with Omniscient Knowledge

In order to detect trends associated with global warming, the control run can be picked from the CMIP6 archive (pre-industrial runs). To detect trends starting at a particular time (e.g., the recovery of stratospheric ozone since year 2000), a control run starting at that time should instead be used (in that case, a "perpetual year 2000" run).

Climate trends are typically inferred from time series in which known climate oscillations (e.g., El Niño Southern Oscillation, Madden-Julian Oscillation, Quasi-Biennal Oscillation, etc) are first removed. Removal techniques include multiple linear regression with uncorrelated explanatory variables (Wilks, 2011) and dynamical linear modeling (Laine et al., 2014). Regardless of the approach chosen, the removal should be performed on the control run before calculating unforced trend distributions.

At the desired two-sided confidence level c_d ($0 < c_d < 100\%$), our method predicts that a trend of magnitude b emerges from internal variability when it occurs over a period of time long enough that unforced trends over that same time period are smaller than b at least $c = \frac{c_d+100}{2}\%$ of the time. In other words, the ToE is the number of time steps y such that:

$$\mathcal{Q}_y(c) = b \tag{1}$$

where $\mathcal{Q}_y : [0,1] \to \mathbb{R}$ is the quantile function for the distribution of y-step unforced trends. In practice, y is estimated numerically as follows:

1. Set y to an appropriate lower bound (e.g., 2);
2. Determine the reference distribution of y-step unforced trends by resampling;
3. Compare p_c, the c-th percentile ($c = \frac{c_d+100}{2}$) in the reference distribution, to b. If $p_c > b$, increase y by one step;
4. Iterate steps 2-3 until $p_c \leq b$. ToE is the value of y needed to obtain this result.

These steps are illustrated in Figure 1c. Alternatively, the method can be used to calculate b given y; b is then the "smallest detectable trend." We will use this approach to quantify the detection power granted by a record of given length.

2.3 Time of Emergence with Observational Knowledge

ToE derived from omniscient knowledge should be interpreted as an ideal quantity. In reality, observational limitations affect the degree of statistical confidence placed in ToE – and by extension, ToE itself. To account for this, we define the operator $\mathcal{K}(*)$ which converts model quantile functions to observed quantile functions. The terminology $\mathcal{K}$ originates from the first intended application of this method to satellite kernel operators. $\mathcal{K}$ represents the process of resampling the model control run so as to reproduce the characteristics of the observing system of interest. For instance, the control run can be resampled to match the spatial and temporal coverage of an observing system (sparse or missing observing locations, changes in coverage over time, etc), or the quality of a data set (instrumental errors, changes in calibration, orbital drift, data processing such as averaging, gridding, interpolation, etc). This process is akin to observing system experiments (OSEs, see e.g., Gelaro & Zhu, 2009).

Applying the new operator $\mathcal{K}$, equation 1 becomes:

$$\mathcal{K}(\mathcal{Q}_y(c)) = b^* \tag{2}$$

where b^* is the c-th percentile of the distribution of unforced trends as would be seen by observations, and can be interpreted as an "observation-equivalent" of b. Should the observing system underestimate the true internal variability, the observed distribution of unforced trends is narrower than its model counterpart, i.e. $|\mathcal{K}(\mathcal{Q}_y(c))| \leq |\mathcal{Q}_y(c)|$. In that case, $b^* \leq b$ (see Figure 1d), reflecting that the apparent detection power granted by the observational record is inflated. In order to adjust ToE for this effect, we calculate the observation-equivalent number of time steps y^* after which the model predicts that the observation-equivalent trend b^* emerges at confidence level c_d:

$$y^* \ \mid \mathcal{Q}_{y^*}(c) = b^* \tag{3}$$

Generally speaking, $|\mathcal{Q}_y(c)|$ decreases when y increases. Thus, when $b^* < b$ then $y^* > y$: it takes longer for a trend to emerge in the observational record when accounting for a variability deficit (see Figure 1e).

In specific cases, observing systems can also overestimate variability. For instance, satellite retrieval methods that rely on averaging (or smoothing) kernels redistribute variability between levels/grid cells, occasionally producing anomalously high local variability. In that case, the method described above predicts that $b^* > b$, and therefore, $y^* < y$: the ToE estimated from observational knowledge is overestimated, and the adjustment therefore yields a shorter ToE. In the rest of the paper, the confidence level c_d is set to 95% ($c = 97.5$), though we recognize that it is an arbitrary choice. We note that confidence levels are still relevant for mission design, policy development, and decision making (though they should be one aspect of the broader context rather than the sole focus, see e.g. Amrhein et al. (2019)).

3 Validation and Discussion of the Method

3.1 Comparison with Other Methods

As previously discussed, the premise behind the concept of ToE is the statistical agreement about a trend in a time series. Given a large ensemble for such a time series, emergence can be determined empirically as the time beyond which the fraction of ensemble members that predict the same sign change is $\frac{c_d+100}{2}$% or greater. This empirical metric is what L17 capture analytically, and serves as benchmark to validate our new method.

Figure 2a shows that for normally distributed residuals (for the sake of comparison with L17), the new method yields nearly identical results to the other definitions. Some differences attributable to the numerical nature of the new method exist, but they remain small for trends encompassing orders of magnitude relevant to real climate signals (from 0.1% to 10% of the magnitude of the noise per unit time). We conclude that the new method provides accurate estimates of ToE, provided the control run is long enough.

3.2 Precision, Performance, Limitations

Since the ToE adjustment described in Section 2.3 is based on a control simulation, the question of the realism of the control simulation is pertinent. In the context of this study, realism most directly concerns the magnitude of internal variability. The analysis presented in Figure 2b shows that the adjustment to ToE (as a percentage of ToE) and stochastic spread around it are largely unaffected by misrepresentations of the magnitude of internal variability in the control simulation. The outlier at (0.2,28.5) for trend magnitude 0.1 results from the unadjusted ToE being extremely small (and unrealistic): nearly always 7 time steps, with an adjustment of 2 time steps (in percentage, 28.57%). These values lie well outside the range typically relevant to climate studies. Similarly, the stochastic spread increases at the largest values of ToE (>1500 time steps for trend 0.001 and large standard deviation ratios).

Other aspects of the realism of the control run, such as the frequency distribution of its internal variability, are secondary since climate oscillations (peaks in the power spectrum) are removed from the control simulation. That being said, the removal may leave behind residual variability at some frequencies. An analysis similar to that shown in Fig-

a)

Trend [arb. units/time step]	ToE [time steps] Empirical	R24	L17
0.001	356.5	357.2	359.4
0.01	77.2	77.5	78.1
0.1	16.9	17.0	17.6

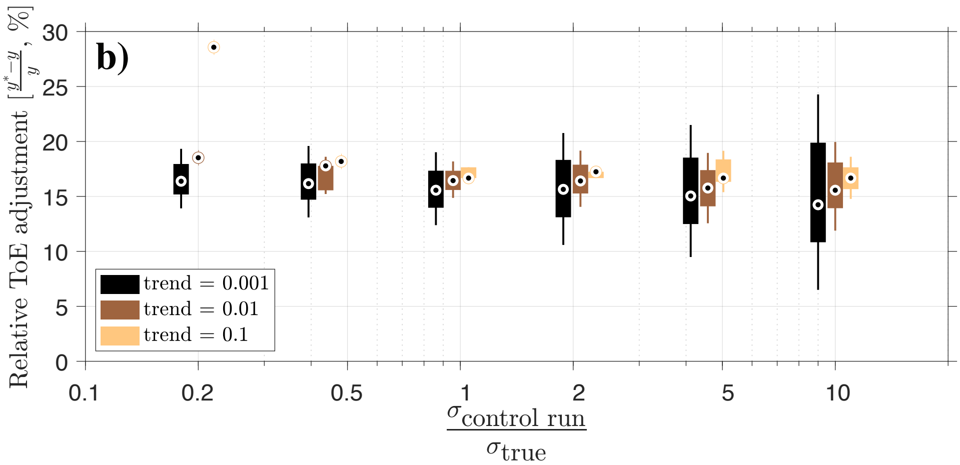

Figure 2. a) Validation of this method (R24) against the empirical definition of ToE and the L17 analytical equivalent, shown as the central estimate of 300 realizations. b) Sensitivity of the ToE adjustment to misrepresentations of the magnitude of internal variability in the control run for 1000 realizations, shown as the ratio of the standard deviation of its residuals to the "true" standard deviation (σ_{true}=1 arbitrary units). Adjustments to ToE are shown for an arbitrary observing system which scales the residuals by 0.8. All time series used are synthetic Gaussian white noise with 1400×12 time steps – matching the length of the Large Ensemble Community Project (LENS2, Rodgers et al., 2021).

ure 2b (not shown) indicates that this should generally not affect the accuracy of the results.

A practical limitation to the new method is its computational cost. Performing linear regressions for all possible y-step periods in a long model run is a time consuming task, especially when it has to be repeated for multiple locations, confidence levels, or trend magnitudes. Implementing the steps in Section 2.2 by dichotomy ensures that, for a control run of length L time steps it always takes $\lfloor \log_2(L) \rfloor$ iterations to calculate ToE.

4 Applications

While long, uninterrupted, and unbiased records are optimal to evaluate climate signals, only few such records exist – even the Keeling Curve experienced minor data gaps, and widely used sea surface temperature measurements have been affectede by changes in data collection techniques (Kent et al., 2010). Nevertheless, the use of existing records for trend analysis remains crucial. This principle extends beyond the historical context: in a theoretical scenario involving solar radiation management, the introduction of aerosols in the stratosphere should be continuously adjusted using observations (MacMartin et al., 2014), in which case integrating observational uncertainties would be important from a policy standpoint. In this section, we discuss the effects of data gaps and data post-processing.

4.1 Effects of Limited Temporal Coverage

Irregularly sampled time series data can yield biased trend estimates and trend uncertainties. The method introduced in this paper can account for the loss of information due to the temporal sampling of an observing system, by using an operator $\mathcal{K}$ (Section 2.3) that samples the control simulation with the same timing as that of the observational record of interest. This way, one ensures that the distribution of unforced trends built using $\mathcal{K}$ captures the effects of temporal sampling. These effects can be large: an application of this method in ongoing research shows that the historic timing of the high-altitude balloon record (Engel et al., 2009) used to evaluate trends in the circulation of the stratosphere is responsible for a 20-year delay in the emergence of potential trends. Another application for the development of the STRAtosphere TO Surface (STRATOS) satellite mission proposal (to study long-term changes in the stratosphere and their impacts at the surface) showed that the proposed accompanying *in situ* validation campaign would still be useful even if the physical recovery of 50% of its balloon-borne measurements failed.

To generalize these results, Figure 3 quantifies the effects of arbitrary data gaps and degraded sampling frequency on the magnitude of the smallest detectable trends (at the 95% confidence level). As one may expect, the detection power stagnates for the duration of data gaps. Once data collection resumes, assuming accurate calibration, the detection power is recovered at a rate that depends on the size and timing of the gap (not shown). In Figure 3, the lower the sampling frequency the lower the detection power. The presence of autocorrelation (memory) aggravates the problem, because autocorrelation diminishes the amount of independent information conveyed by individual data points (and more data must be collected to compensate). This effect becomes less prevalent as the sampling frequency degrades, because data points that are further apart may co-vary less.

4.2 Effects of Data Post-Processing

Measures of internal variability derived from observations are sensitive to data collection and treatment procedures. For instance, averaging and interpolation methods used to convert scattered observations into gridded products have detrimental effects: Hofstra et al. (2010) find systematically misrepresented variance especially in upper percentiles. Relatedly, Lin and Huybers (2019) concluded that changes in spatial sampling must be taken into account when reconciling observed trends with climate projections. Other data reporting issues are relevant to this section, for instance rounding-related errors (Rhines et al., 2015), or the conversion of measured variables into other quantities. For instance, in the conversion of N_2O measurements into age of air using empirical relationships (Boering et al., 1996; Linz et al., 2017), our method showed that seasonal, instead of monthly, measurements are sufficient to preserve the detection power needed to study trends in the age-based Brewer-Dobson circulation (defined by Linz et al., 2016). This result also informed the budget and logistics of STRATOS.

Another example particularly relevant to remote sensing is the use of smoothing kernels (or averaging kernels) in satellite retrieval algorithms. Such algorithms generally estimate a quantity as the weighted average of neighboring observations, often incorporating *a priori* climatological information. This kind of averaging affects both variability and trends. If the averaging kernels and a detailed description of the retrieval algorithm are available, it is possible to define the operator $\mathcal{K}$ to sample the control run as the observing system does the real world. Figure 4a shows this process for a simple

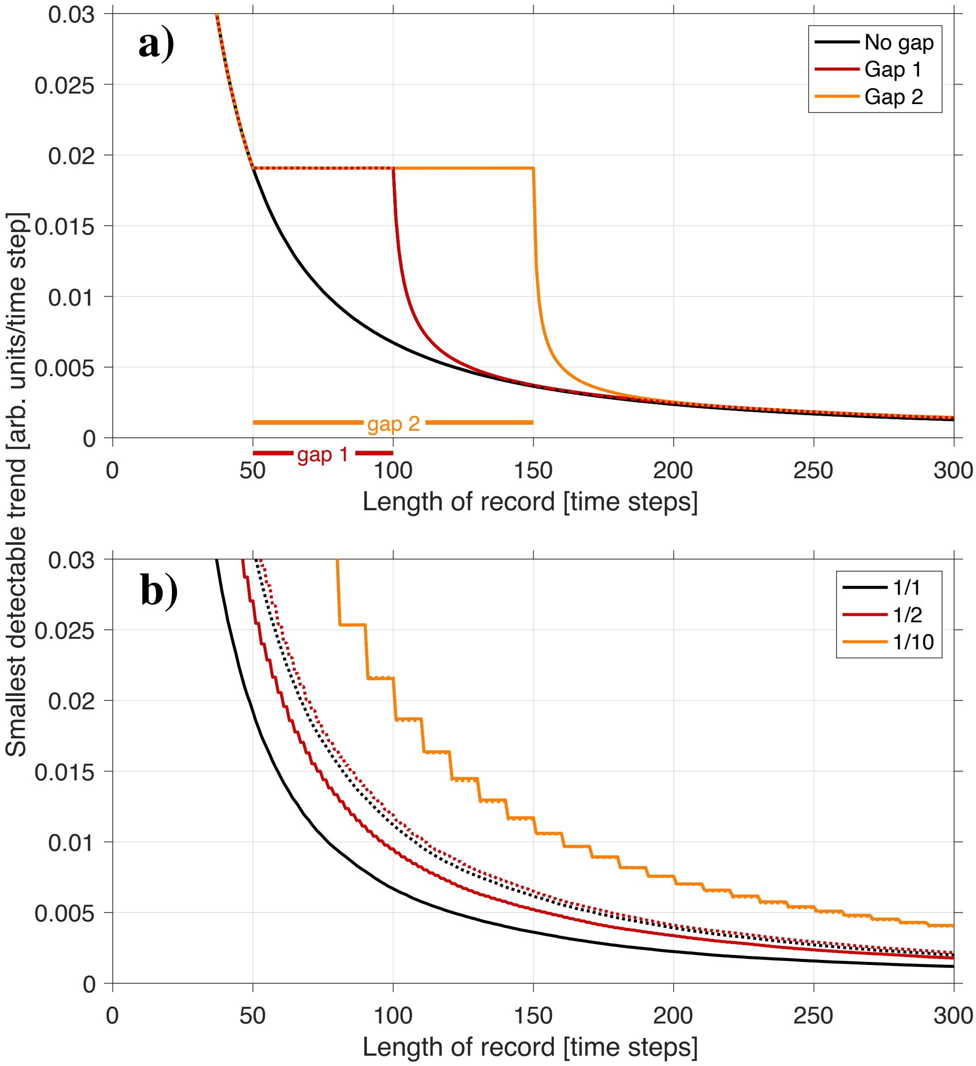

Figure 3. Loss of detection power due to a) data gaps and b) varying sampling frequencies. Sampling frequency is relative to an optimal frequency of 1 (e.g., 1/2 indicates that every other time step is sampled). Frequencies greater (smaller) than 1 indicate redundancy (sparsity) in the data. Solid lines denote zero lag-one autocorrelation (acf(1)=0) and dashed lines denote acf(1)=0.5.

hypothetical example, and Figure 4b shows that a significant adjustment to ToE (20-40%) is necessary. As a concrete example, uncertainty in satellite-derived trends in the midlatitude ozone layer is partly attributable to differences between satellite platforms (Ball et al., 2019). An application of our method (not shown) showed that the smoothing kernels in solar backscatter ultraviolet retrievals (see Kramarova et al., 2013) considerably reduce the degree of confidence in ozone trends, highlighting the potential for complications in direct comparison between products (see Godin-Beekmann et al., 2022). Such complications also concern the production of merged satellite products; correction schemes are typically applied to minimize differences between platforms, but residual errors can be difficult to assess in the presence of natural variability (Randel, 2010), affecting trend estimates and our understanding of internal variability (CCMVal, 2010; Randel, 2010).

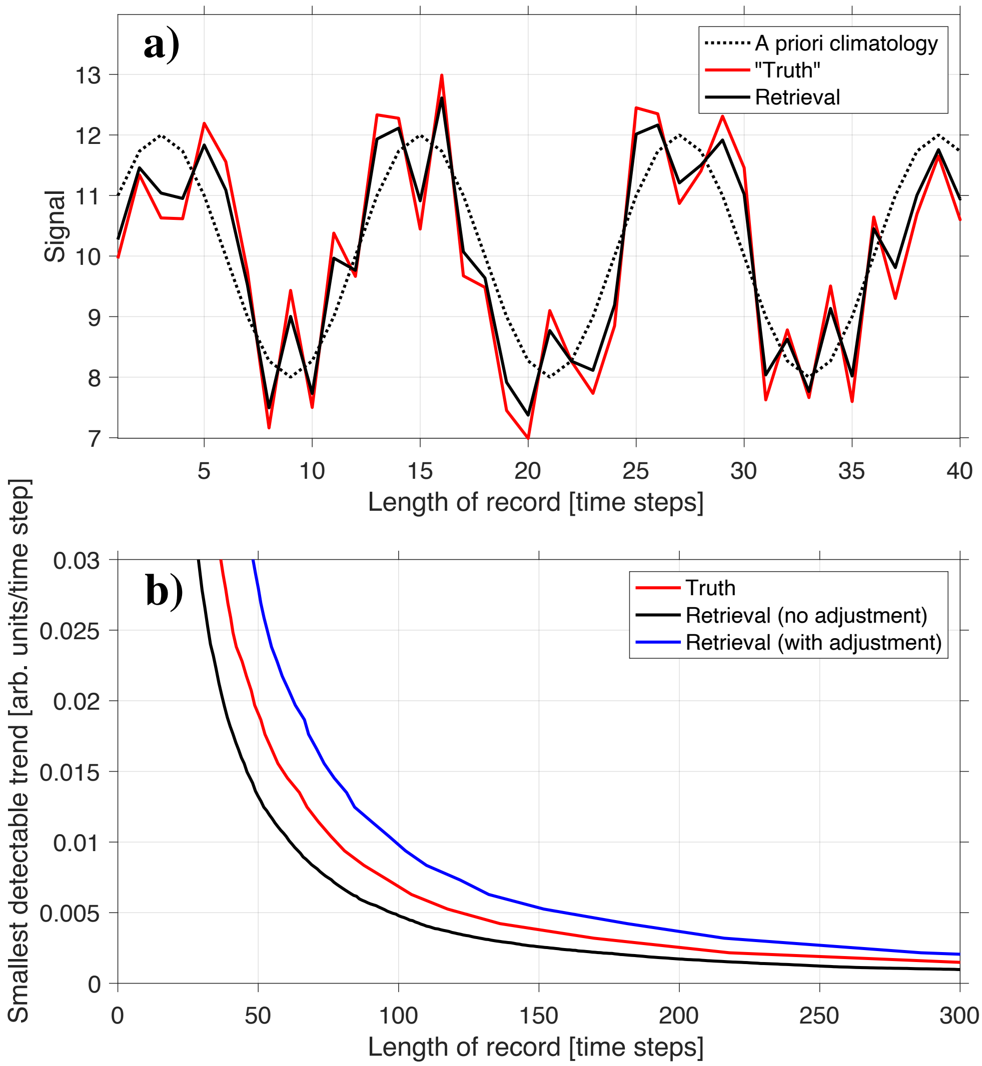

Figure 4. a) Seasonally-varying synthetic time series showing the effect of a retrieval algorithm using *a priori* information and observational data in a 30%:70% ratio. b) Loss of detection power due to the influence of the *a priori* information. The red line is the detection power with perfect knowledge when imposing a chosen trend onto the "truth" in a), the black and blue lines are for observational knowledge without and with adjustment for the retrieval algorithm, respectively.

5 Conclusions

We introduce a new definition for the time of emergence (ToE) of forced climate signals, based on the resampling of a control climate simulation. Results compare well with the definition from L17 and with an empirical definition. Our definition eliminates the key assumption that the "residuals" representing internal variability be normally distributed about the trend of interest.

Further, the new method can adjust ToE to account for the limitations of observing systems, to systematically handle data of varying quality in climate records. We find that the relative adjustment to ToE is accurate even when using a control simulation that under or overestimates internal variability (though ToE itself is inaccurate in that case).

Lastly, the new method serves as a quantitative tool to guide the development of future observing platforms and mitigation strategies: by taking into consideration scientific aspects within the framework of budgetary and logistical constraints, one can assess the cost and technical feasibility of future observing systems (National Academies of Sciences & Medicine, 2018) to ensure that observational priorities are aligned with future scientific and societal needs.

6 Open Research

No actual measurements or data sources were used in this manuscript; the data are synthetic in nature and produced by random number generation (process described in Section 2.1).

Acknowledgments
LR was funded by NASA NIP 80NSSC21K0943 and the William F. Milton Fund, ML was funded by NASA NIP 80NSSC21K0943 and NASA ACMAP 80NSSC23K1005, JL was funded by NSF 2023483. The authors report no real or perceived financial conflicts of interests. The computations in this study were run on the FASRC Cannon cluster at Harvard University. We thank Karen McKinnon, Aaron Match, and Todd Mooring for helpful discussions during the preparation of the manuscript.

References

Abalos, M., Calvo, N., Benito-Barca, S., Garny, H., Hardiman, S. C., Lin, P., ... others (2021). The brewer–dobson circulation in cmip6. *Atmospheric Chemistry and Physics*, *21*(17), 13571–13591.

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305–307.

Ball, W. T., Alsing, J., Staehelin, J., Davis, S. M., Froidevaux, L., & Peter, T. (2019). Stratospheric ozone trends for 1985–2018: sensitivity to recent large variability. *Atmospheric Chemistry and Physics*, *19*(19), 12731–12748.

Barnes, E. A., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., & Anderson, D. (2019). Viewing forced climate patterns through an ai lens. *Geophysical Research Letters*, *46*(22), 13389–13398.

Barnett, T. P., & Schlesinger, M. E. (1987). Detecting changes in global climate induced by greenhouse gases. *Journal of Geophysical Research: Atmospheres*, *92*(D12), 14772–14780.

Biesbroek, G. R., Klostermann, J. E., Termeer, C. J., & Kabat, P. (2013). On the nature of barriers to climate change adaptation. *Regional Environmental Change*, *13*, 1119–1129.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., ... others (2014). *Detection and attribution of climate change: from global*

to regional.
Boering, K. A., Wofsy, S., Daube, B., Schneider, H., Loewenstein, M., Podolske, J., & Conway, T. (1996). Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science*, *274*(5291), 1340–1343.
Bothe, O., Jungclaus, J. H., & Zanchettin, D. (2013). Consistency of the multi-model cmip5/pmip3-past1000 ensemble. *Climate of the Past*, *9*(6), 2471–2487.
Branstator, G., & Selten, F. (2009). "modes of variability" and climate change. *Journal of Climate*, *22*(10), 2639–2658.
CCMVal, S. (2010). Sparc report on the evaluation of chemistry-climate models. *(No Title)*.
Christensen, J. H., Hewitson, B., Busuioc, A., et al. (2007). *Regional climate projections. in 'climate change 2007: the physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change'.(eds s solomon, d qin, m manning, z chen, m marquis, kb averyt, m tignor, hl miller) pp. 847–940.* Cambridge University Press: Cambridge, UK.
Collins, M., AchutaRao, K., Ashok, K., Bhandari, S., Mitra, A. K., Prakash, S., ... Turner, A. (2013). Observational challenges in evaluating climate models. *Nature Climate Change*, *3*(11), 940–941.
Collins, M., Osborn, T. J., Tett, S. F., Briffa, K. R., & Schweingruber, F. H. (2002). A comparison of the variability of a climate model with paleotemperature estimates from a network of tree-ring densities. *Journal of Climate*, *15*(13), 1497–1515.
Danabasoglu, G. (2008). On multidecadal variability of the atlantic meridional overturning circulation in the community climate system model version 3. *Journal of Climate*, *21*(21), 5524–5544.
Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, *38*, 527–546.
Diffenbaugh, N. S., & Scherer, M. (2011). Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries: A letter. *Climatic Change*, *107*, 615–624.
Drijfhout, S., Hazeleger, W., Selten, F., & Haarsma, R. (2008). Future changes in internal variability of the atlantic meridional overturning circulation. *Climate dynamics*, *30*, 407–419.
Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., ... others (2009). Age of stratospheric air unchanged within uncertainties over the past 30 years. *Nature Geoscience*, *2*(1), 28–31.
Gelaro, R., & Zhu, Y. (2009). Examination of observation impacts derived from observing system experiments (oses) and adjoint models. *Tellus A*, *61*(2), 179–193.
Gessner, C., Fischer, E. M., Beyerle, U., & Knutti, R. (2021). Very rare heat extremes: quantifying and understanding using ensemble reinitialization. *Journal of Climate*, *34*(16), 6619–6634.
Giorgi, F., & Bi, X. (2009). Time of emergence (toe) of ghg-forced precipitation change hot-spots. *Geophysical Research Letters*, *36*(6).
Godin-Beekmann, S., Azouz, N., Sofieva, V. F., Hubert, D., Petropavlovskikh, I., Effertz, P., ... others (2022). Updated trends of the stratospheric ozone vertical distribution in the 60° s–60° n latitude range based on the lotus regression model. *Atmospheric Chemistry and Physics*, *22*(17), 11657–11673.
Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*(1).
Hegerl, G. C., Hasselmann, K., Cubasch, U., Mitchell, J. F., Roeckner, E., Voss, R., & Waszkewitz, J. (1997). Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change.

Climate Dynamics, *13*, 613–634.

Hegerl, G. C., voN SToRcH, H., Hasselmann, K., Santer, B. D., Cubasch, U., & Jones, P. D. (1996). Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *Journal of Climate*, *9*(10), 2281–2306.

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate dynamics*, *35*, 841–858.

Islam, N., & Winkel, J. (2017). Climate change and social inequality.

Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., ... Turkington, T. (2023). Importance of internal variability for climate model assessment. *npj Climate and Atmospheric Science*, *6*(1), 68.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2* (Vol. 289). John wiley & sons.

Kent, E. C., Kennedy, J. J., Berry, D. I., & Smith, R. O. (2010). Effects of instrumentation changes on sea surface temperature measured in situ. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 718–728.

Kim, W. M., Yeager, S., Chang, P., & Danabasoglu, G. (2018). Low-frequency north atlantic climate variability in the community earth system model large ensemble. *Journal of Climate*, *31*(2), 787–813.

Kramarova, N., Bhartia, P. K., Frith, S., McPeters, R., & Stolarski, R. (2013). Interpreting sbuv smoothing errors: An example using the quasi-biennial oscillation. *Atmospheric Measurement Techniques*, *6*(8), 2089–2099.

Laepple, T., Ziegler, E., Weitzel, N., Hébert, R., Ellerhoff, B., Schoch, P., ... others (2023). Regional but not global temperature variability underestimated by climate models at supradecadal timescales. *Nature Geoscience*, 1–9.

Laine, M., Latva-Pukkila, N., & Kyrölä, E. (2014). Analysing time-varying trends in stratospheric ozone time series using the state space approach. *Atmospheric Chemistry and Physics*, *14*(18), 9707–9725.

Lee, D., Min, S.-K., Park, C., Suh, M.-S., Ahn, J.-B., Cha, D.-H., ... Kang, H.-S. (2016). Time of emergence of anthropogenic warming signals in the northeast asia assessed from multi-regional climate models. *Asia-Pacific Journal of Atmospheric Sciences*, *52*, 129–137.

Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., ... others (2023). Climate change 2023: synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change.

Li, J., Thompson, D. W., Barnes, E. A., & Solomon, S. (2017). Quantifying the lead time required for a linear trend to emerge from natural climate variability. *Journal of Climate*, *30*(24), 10179–10191.

Lin, M., & Huybers, P. (2019). If rain falls in india and no one reports it, are historical trends in monsoon extremes biased? *Geophysical Research Letters*, *46*(3), 1681–1689.

Linz, M., Plumb, R. A., Gerber, E. P., Haenel, F. J., Stiller, G., Kinnison, D. E., ... Neu, J. L. (2017). The strength of the meridional overturning circulation of the stratosphere. *Nature Geoscience*, *10*(9), 663–667.

Linz, M., Plumb, R. A., Gerber, E. P., & Sheshadri, A. (2016). The relationship between age of air and the diabatic circulation of the stratosphere. *Journal of the Atmospheric Sciences*, *73*(11), 4507–4518.

Lyu, K., Zhang, X., Church, J. A., Slangen, A. B., & Hu, J. (2014). Time of emergence for regional sea-level change. *Nature Climate Change*, *4*(11), 1006–1010.

MacMartin, D. G., Kravitz, B., Keith, D. W., & Jarvis, A. (2014). Dynamics of the coupled human–climate system resulting from closed-loop control of solar geoengineering. *Climate dynamics*, *43*, 243–258.

Madden, R. A., & Ramanathan, V. (1980). Detecting climate change due to increasing carbon dioxide. *Science*, *209*(4458), 763–768.

Mahlstein, I., Knutti, R., Solomon, S., & Portmann, R. W. (2011). Early onset of significant local warming in low latitude countries. *Environmental Research Letters*, *6*(3), 034009.

Maraun, D. (2013). When will trends in european mean and heavy daily precipitation emerge? *Environmental Research Letters*, *8*(1), 014004.

McKinnon, K. A., & Deser, C. (2021). The inherent uncertainty of precipitation variability, trends, and extremes due to internal variability, with implications for western us water resources. *Journal of Climate*, *34*(24), 9605–9622.

McKinnon, K. A., & Simpson, I. R. (2022). How unexpected was the 2021 pacific northwest heatwave? *Geophysical Research Letters*, *49*(18), e2022GL100380.

Mitchell, D., Thorne, P., Stott, P., & Gray, L. (2013). Revisiting the controversial issue of tropical tropospheric temperature trends. *Geophysical Research Letters*, *40*(11), 2801–2806.

National Academies of Sciences, E., & Medicine. (2018). *Thriving on our changing planet: A decadal strategy for earth observation from space.* National Academies Press.

Nguyen, T.-H., Min, S.-K., Paik, S., & Lee, D. (2018). Time of emergence in regional precipitation changes: an updated assessment using the cmip5 multi-model ensemble. *Climate Dynamics*, *51*, 3179–3193.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, *185*, 71–110.

Pörtner, H., Roberts, D. C., Parmesan, C., Adams, H., Adelekan, I., Adler, C., ... others (2023). *Ipcc 2022: Technical summary, working group ii impacts, adaptation and vulnerability* (Unpublished doctoral dissertation). Intergovernmental Panel on Climate Change.

Qasmi, S., & Ribes, A. (2022). Reducing uncertainty in local temperature projections. *Science Advances*, *8*(41), eabo6872.

Randel, W. J. (2010). Variability and trends in stratospheric temperature and water vapor. *The Stratosphere: Dynamics, Transport, and Chemistry*, *190*, 123–135.

Rhines, A., Tingley, M. P., McKinnon, K. A., & Huybers, P. (2015). Decoding the precision of historical temperature observations. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 2923–2933.

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... others (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, *12*(4), 1393–1411.

Santer, B. D., Mikolajewicz, U., Brüggemann, W., Cubasch, U., Hasselmann, K., Höck, H., ... Wigley, T. M. (1995). Ocean variability and its influence on the detectability of greenhouse warming signals. *Journal of Geophysical Research: Oceans*, *100*(C6), 10693–10725.

Schipper, E. L. F. (2020). Maladaptation: when adaptation to climate change goes very wrong. *One Earth*, *3*(4), 409–414.

Schmidt, G. A., Jones, G. S., & Kennedy, J. J. (2023). Comment on “advanced testing of low, medium, and high ecs cmip6 gcm simulations versus era5-t2m” by n. scafetta (2022). *Geophysical Research Letters*, *50*(18), e2022GL102530.

Shepherd, T. G. (2021). Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, *169*(1-2), 2.

Sui, Y., Lang, X., & Jiang, D. (2014). Time of emergence of climate signals over china under the rcp4. 5 scenario. *Climatic change*, *125*, 265–276.

Tao, M., Konopka, P., Wright, J. S., Liu, Y., Bian, J., Davis, S. M., ... Ploeger, F. (2023). Multi-decadal variability controls short-term stratospheric water vapor trends. *Communications Earth & Environment*, *4*(1), 441. Retrieved from `https://doi.org/10.1038/s43247-023-01094-9` doi: 10.1038/s43247-023-01094-9

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: math-*

ematical, physical and engineering sciences, *365*(1857), 2053–2075.

Thompson, D. W., Barnes, E. A., Deser, C., Foust, W. E., & Phillips, A. S. (2015). Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, *28*(16), 6443–6456.

Wigley, T., & Jones, P. (1981). Detecting co2-induced climatic change. *Nature*, *292*(5820), 205–208.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Williamson, M. S., Thackeray, C. W., Cox, P. M., Hall, A., Huntingford, C., & Nijsse, F. J. (2021). Emergent constraints on climate sensitivities. *Reviews of Modern Physics*, *93*(2), 025004.

Wills, R. C., Battisti, D. S., Armour, K. C., Schneider, T., & Deser, C. (2020). Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *Journal of Climate*, *33*(20), 8693–8719.

Wu, Y., Miao, C., Fan, X., Gou, J., Zhang, Q., & Zheng, H. (2022). Quantifying the uncertainty sources of future climate projections and narrowing uncertainties with bias correction techniques. *Earth's Future*, *10*(11), e2022EF002963.

Zappa, G., Hoskins, B. J., & Shepherd, T. G. (2015). Improving climate change detection through optimal seasonal averaging: The case of the north atlantic jet and european precipitation. *Journal of Climate*, *28*(16), 6381–6397.

Zelle, H., van Oldenborgh, G. J., Burgers, G., & Dijkstra, H. (2005). El niño and greenhouse warming: Results from ensemble simulations with the ncar ccsm. *Journal of Climate*, *18*(22), 4669–4683.