AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL074596

Key Points:
- The 10,000-member ensemble sampling observation and coverage uncertainty in surface temperatures created by multiresolution lattice kriging
- This allows probabilistic analysis of past regional temperatures and their trends
- Prior to the 1920s, a confident identification of the El Nino-Southern Oscillation state may not be possible

Correspondence to:
M. Ilyas,
maryam.ilyas.14@ucl.ac.uk

Citation:
Ilyas, M., C. M. Brierley, and S. Guillas (2017), Uncertainty in regional temperatures inferred from sparse global observations: Application to a probabilistic classification of El Niño, *Geophys. Res. Lett.*, *44*, 9068–9074, doi:10.1002/2017GL074596.

Received 20 JUN 2017
Accepted 16 AUG 2017
Accepted article online 21 AUG 2017
Published online 9 SEP 2017

©2017. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Uncertainty in regional temperatures inferred from sparse global observations: Application to a probabilistic classification of El Niño

Maryam Ilyas[1], Christopher M. Brierley[2], and Serge Guillas[1]

[1]Department of Statistical Science, University College London, London, UK, [2]Department of Geography, University College London, London, UK

Abstract Instrumental records showing increases in surface temperature are some of the robust and iconic evidence of climate change. But how much should we trust regional temperature estimates interpolated from sparse observations? Here we quantify the uncertainty in the instrumental record by applying multiresolution lattice kriging, a recently developed interpolation technique that leverages the multiple spatial scales of temperature anomalies. The probability of monthly anomalies across the globe is represented by an ensemble, based on HadCRUT4 and accounting for observational and coverage uncertainties. To demonstrate the potential of these new data, we investigate the area-averaged temperature anomalies over the Niño 3.4 region in the equatorial Pacific. Having developed a definition of the El Niño–Southern Oscillation (ENSO) able to cope with probability distribution functions, we classify the ENSO state for each year since 1851. We find that for many years it is ambiguous as to whether there was an El Niño or not from the Niño 3.4 region alone. These years are mainly before 1920, but also just after World War II.

1. Introduction

Long-term increases in temperatures across the globe are one of the key indicators of climate change. Measurements of air temperature from ground-based stations and sea surface temperature (SST) from ships and buoys in the oceans are combined to generate global surface temperature estimates. Initially, these raw data undergo a quality control process to remove random errors (i.e, observer or instrumental errors). Homogeneity assessment is further carried out to eliminate systematic biases (i.e., nonclimatic signals). Multiple independent data sets have been compiled that treat these errors in different ways. They are commonly better known by their acronyms: MLOST by *Smith et al.* [2008], GISS by *Hansen et al.* [2010], JMA by *Ishii et al.* [2005], HadCRUT by *Morice et al.* [2012], and BEST by *Rohde et al.* [2013].

In addition to quality control and data homogenization, the deficient spatial coverage has been highlighted as an important issue. It can, for example, lead to an underestimation of the global mean temperature trend by biased sampling of the Arctic [*Cowtan and Way*, 2014] which is warming particularly quickly. Here we investigate and quantify the impact of coverage uncertainties using HadCRUT4 [*Morice et al.*, 2012] which clearly flags locations where reliable observations are missing (section 2.1). In section 2.2, we apply a novel multiresolution interpolation technique [*Nychka et al.*, 2015] to create a spatially complete ensemble of equally probable realizations of temperatures (section 3). The quantified uncertainties in the resulting ensemble are evaluated against published estimates at the global mean scale (section 3.1). As expected, the uncertainties arising from poorer coverage tend to be greater at the local scale. We illustrate this by looking at the El Niño–Southern Oscillation (ENSO) in section 4.2, using an enhanced definition of El Niño designed for probability density functions (section 4.1).

2. Methods

2.1. Data

HadCRUT4 is chosen for this exploration into the impact of spatial uncertainties, because it comes with a quantification of its observational uncertainties [*Morice et al.*, 2012]. HadCRUT4 temperature anomalies (specifically version 4.5.0.0) are a global combination of land surface air temperature (CRUTEM4) and sea surface monthly temperature anomalies (HadSST3). HadCRUT4 consists of 100 ensemble members that characterize observational uncertainties arising from the nonclimatic factors [*Morice et al.*, 2012]. These factors include

missing observations that deteriorate the computation of a station's monthly average and climatology, as well as changes in station properties that are left uncorrected (equipment, location, measurement time, and urban heat effects). The spatial coverage of the data varies with the available observations. Even its most complete coverage presents gaps over Africa and the polar regions [*Cowtan and Way*, 2014].

HadCRUT4 flags any 5° latitude by 5° longitude grid box as missing data, when there are no known observations for that month [*Morice et al.*, 2012]. This in contrast to other databases that use various methods of spatial interpolation to fill in missing grid boxes. MLOST uses linear spatial interpolation using nearby stations [*Smith et al.*, 2008]. GISS incorporates a linear inverse distance weighting to use information from stations up to 1200 km of the prediction location [*Hansen et al.*, 2010]. Kriging can also be used to fill any data gaps based upon either covariances (JMA) [*Ishii et al.*, 2005] or correlations (BEST) [*Rohde et al.*, 2013]. While all these interpolated data sets provide error estimates, they do not benefit from the recent leaps forward made in approaches to quantify uncertainty in interpolations.

2.2. Statistical Approach

Geophysical variables typically exhibit heterogeneous variations at a range of scales [*Berliner et al.*, 2012]. Standard covariance functions used in kriging can often oversimplify the complicated spatial dependence prevalent in variables such as temperature. Moreover, it is challenging for covariance functions with light tails to capture variation among observations widely separated over the spatial domain (e.g., Matérn) [*Kleiber and Porcu*, 2015] as it is the case for temperatures over the globe. Newly developed statistical techniques outperform covariance-based kriging in terms of accuracy and uncertainty quantification over the globe [*Chang et al.*, 2015].

Multiresolution lattice kriging (MRLK) [*Nychka et al.*, 2015] is designed to predict sparsely observed two-dimensional spatial fields. The approach's advantage comes from its use of multiple resolutions to represent spatial dependence at various levels (local and regional). Additionally, it uses sparse approximations of the covariance matrix that ensures computational efficiency. Recent revisions mean that the routine incorporates full spherical domains (https://cran.r-project.org/web/packages/LatticeKrig). At each resolution level, it organizes the nodes (centers) of compactly supported radial basis functions on a geodesic grid of points using Wendland covariance functions and the great circle distance. The geodesic grid is used instead of a rectangular grid to avoid distortions at the poles. The centers of basis functions are chosen by considering points on a regular icosahedron. The triangles are further subdivided four times meaning that the number of basis functions at each resolution level roughly increases by a factor of 4.

We apply MRLK to the HadCRUT4 data set with lattices at three resolutions (Figure 1b), so that the number of basis functions (3366) is greater than the number of grid boxes (2592). This enables us to capture variations from global to regional scales [*Katzfuss*, 2017]. A common value (8.95) of spatial autoregression weight (i.e., a measure of the strength of random variations across the field) was used at all levels. The proportion of global variance contained at each resolution (i.e., the distribution of the total variability) are $\alpha = (0.211, 0.6703, \text{and } 0.1187)^T$ (i.e., two thirds of the temperature variability occur at scales of around 1000 km). These ratios, as well as the spatial autoregression weights, are determined using the least sparse temperature field (February 1988) since they model underlying physical characteristics common across fields. The spatial prediction also requires the estimation of covariance parameters [*Nychka et al.*, 2015]. These are selected with a profile maximum likelihood method for each spatial field since these parameters will vary across months due to short- and long-term variability in spatial patterns of temperatures. Note that we do not consider uncertainties in the model parameters; accounting for them could lead to less confidence in our temperature estimates. This approach allows for the statistical properties to change across the grid, say, with latitude [*Nychka et al.*, 2015].

As an illustration, we present results for May 1861 (Figure 1). This is the month in HadCRUT4 with the minimum spatial coverage, less than 10% (Figure 1a). The Wendland compactly supported covariance function is used to construct radial basis functions at multiple levels (Figure 1b). Using the multiresolution basis functions and observed spatial points, the temperatures are extended to the whole globe (Figure 1c). The pointwise prediction uncertainties arising can be computed (Figure 1d). These uncertainties may be large (over 1°C) far away from the observations but naturally reduce near them. One interesting feature of MRLK is that it "adjusts" individual observations to fit the wider spatial field (cf. Figures 1a and 1c). Inspection of the distribution of these adjustments demonstrates the appropriateness of the prediction uncertainties (Figure 1d). Figure 1e displays the distribution of the deviations of the predicted values from the observed values since MRLK is not an interpolatory scheme. These (standardized) deviations follow a standard normal distribution highlighting the ability of MRLK to make meaningful predictions where data are present.

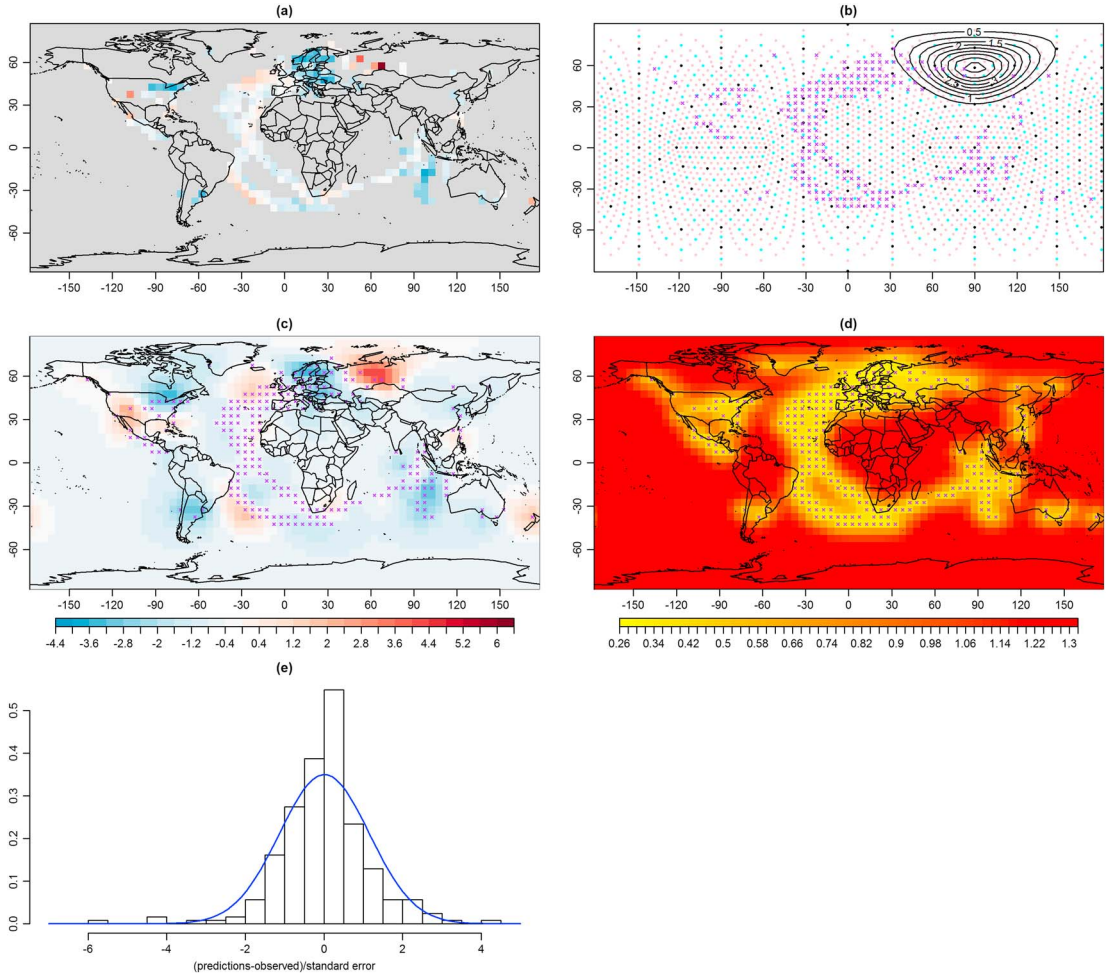

Figure 1. Multiresolution lattice kriging temperature reconstruction and spatial uncertainties. (a) Median of the instrumental reconstructions of the temperature anomalies for May 1861 [*Morice et al.*, 2012]. (b) Location of observations (purple) and the centers of basis functions at the three levels: coarse (black), synoptic (cyan), and fine (pink). Number of basis functions are 162, 642, and 2562, respectively. For illustration, the basis function at a single, coarse node point is shown as contour lines. (c) The spatially complete field of temperatures predicted by multiresolution lattice kriging. (d) Associated uncertainties in those predictions. (e) Distribution of differences between predictions and observations (left in) at the same location in units of one standard deviation of the field's inferred uncertainty.

3. An Ensemble Sampling Uncertainty in Surface Temperature

A 10,000-member ensemble is created to sample both the spatial and observational uncertainties. The ensemble is created via the 100 ensemble members of the sparse HadCRUT4 data set, which sample solely the observational uncertainties [*Morice et al.*, 2012]. For each of these original sparse members, 100 samples are generated from the MRLK's own prediction uncertainty. Rather than sampling each grid point's prediction independently, we instead sample given fixed basis functions. This approach draws strength from the multiresolution aspect of the lattice kriging to effectively represent both large- and small-scale correlation structures in the climate system. Each month is treated independently for consistency with *Morice et al.* [2012].

Additionally, a further 100 members are created by sampling the uncertainty in the MRLK-interpolated median of the original sparse observations [*Morice et al.*, 2012]. These can compared to the initial 100 ensemble members (only characterizing observational uncertainties with spatial gaps) to demonstrate the relative impact of coverage and observational uncertainties.

3.1. Uncertainties in Global Mean Temperature

The importance of the sparsity of the instrumental observations has long been recognized, and the uncertainty caused by it has been estimated before [*Morice et al.*, 2012]. The error arising in an estimate of a global (or regional) mean temperature anomaly can be estimated by subsampling a spatially complete data set

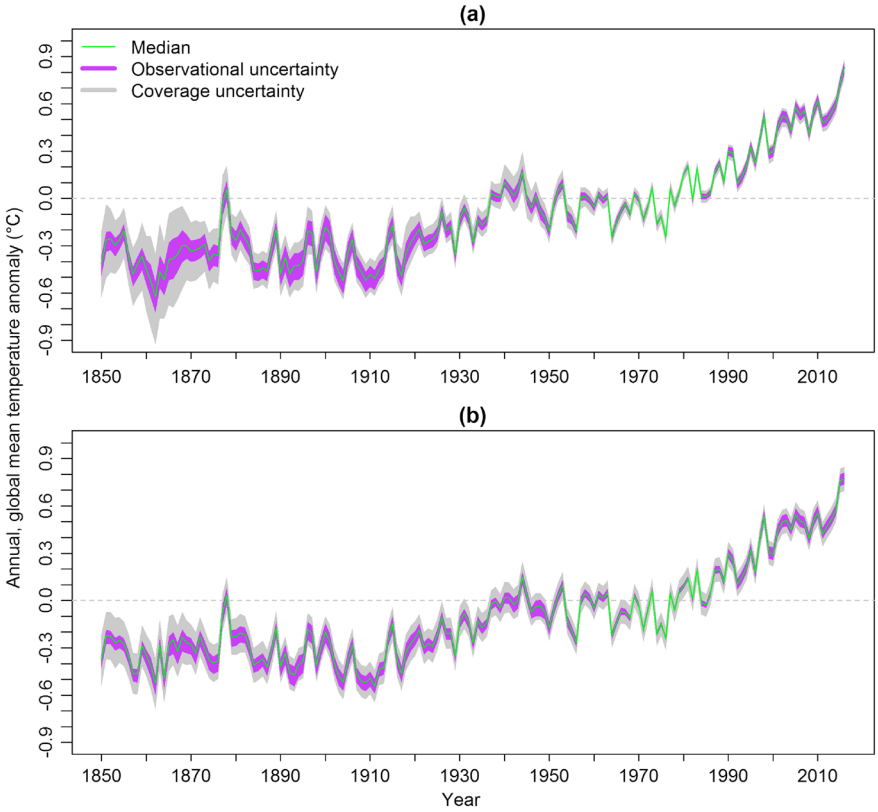

Figure 2. Global mean, annual average temperature anomalies: calculated using (a) the method described here and (b) by *Morice et al.* [2012]. The shaded areas represent the 95% confidence interval estimates associated with coverage uncertainty (gray) and observational uncertainty (purple).

(such as a reanalysis) at the actual, sparse coverage [*Brohan et al.*, 2006]. Applying that approach to the HadCRUT4 data set, *Morice et al.* [2012] were able to estimate the relative importance of coverage uncertainty compared to observational uncertainty at the global and hemispheric level (Figure 2b). As a test of our methodology and MRLK prediction uncertainties, we evaluate those same relative contributions in the global mean temperatures of the new ensemble (Figure 2a). For each year, the median member along with the 95% confidence interval from the two uncertainty estimation approaches is presented. The approach we have adopted (section 3) assumes that uncertainty within the individual observations is correctly sampled by *Morice et al.* [2012].

It is evident that uncertainty arising from the incomplete spatial coverage is substantially larger than that arising solely from the observational errors (Figure 2). This dominance is especially clear prior to the organized global monitoring and collection of weather data due to the establishment of the World Meteorological Organization in the 1950s. Both methodological approaches result in similar estimates of the respective uncertainty contributions. Unsurprisingly, when the uncertainty itself is larger, the impact of the uncertainty quantification approach also becomes larger. The expanded uncertainty estimates in the early record from the new approach (Figure 2a) are perhaps understandable given that it cannot incorporate the physical covariance structures sampled in the reanalysis-based estimate [*Brohan et al.*, 2006]. Nonetheless, we consider the comparison of the two uncertainty estimates to show that our approach is not unreasonable. In fact, it has the advantage of quantifying the uncertainty at more local scales, as will be demonstrated via application to a commonly analyzed region.

4. Niño 3.4 Temperature Anomalies

To demonstrate the benefits of the multiresolution lattice kriging ensemble, we investigate its impact on reconstructions of the Niño 3.4 anomalies. The El Niño–Southern Oscillation (ENSO) is the dominant mode of interannual variability in the global climate, with an El Niño being the warm phase of the oscillation

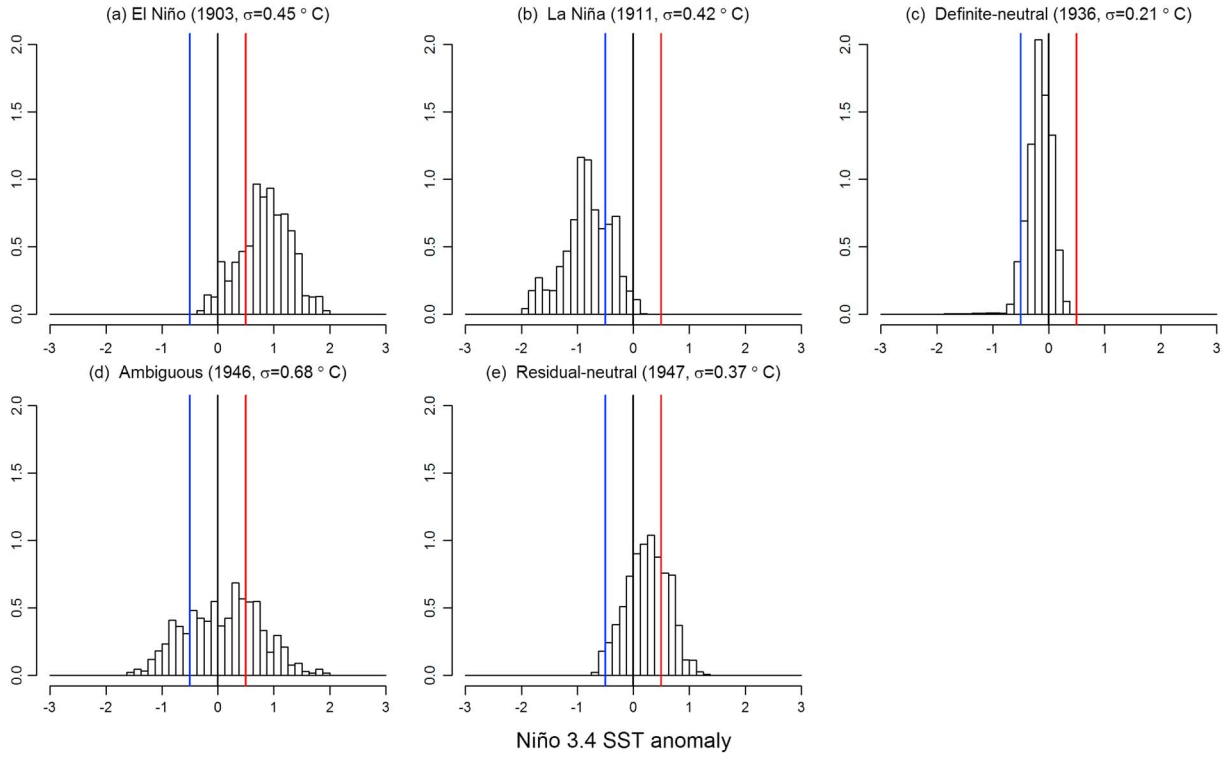

Figure 3. (a–e) The five different states of the El Niño–Southern Oscillation (equation (1) demonstrated using the December-January-February temperature anomaly in the Niño 3.4 region.

[*Trenberth*, 1997]. Of its many definitions, the most common is based on sea surface temperature anomalies in the Niño 3.4 region of the equatorial Pacific (5°S–5°N, 120°W–170°W). Therefore, the reconstruction of spatial surface temperature over this region is of particular interest. Although sea level pressure-derived indices exist [*Ropelewski and Jones*, 1987], SST-based indices are still commonly used [e.g., *Banholzer and Donner*, 2014; *Hartmann et al.*, 2013]. Before the state of ENSO within the ensemble can be identified, we must first revisit how an El Niño is defined.

4.1. Probabilistic El Niño Definition

The traditional definition of ENSO is deterministic, as it requires knowledge of whether the temperature at a time, T_i, exceeds a critical value, T_{crit}, often 0.5°C [*Trenberth*, 1997] as used here. Temperatures greater than this threshold are termed "El Niños" and less than $-T_{\text{crit}}$ are "La Niñas." ENSO is considered to be in a "neutral" state otherwise. The temperature anomalies in the Niño 3.4 region are often computed with respect to a 30 year moving climatology (updated every 5 years) [*Lindsey*, 2013]. This traditional definition has not been designed to handle uncertain temperatures. We therefore introduce a more sophisticated definition that can cope with temperature probability density functions (PDF).

Similar to the usual definition, our definition (equation (1) is also based on a critical temperature, T_{crit}, but additionally a critical probability, P_{crit} (taken as 0.66 here, following the "likely" criterion of *Mastrandrea et al.* [2010]). It is designed such that as the spread in each temperature observation (σ_i) collapses to 0, the traditional definition emerges. Examples of the classification in action are shown in Figure 3.

$$\text{ENSO}_i = \begin{cases} \text{El Niño} & \text{if } p(T_i \geq T_{\text{crit}}) \geq P_{\text{crit}} \\ \text{La Niña} & \text{if } p(T_i \leq -T_{\text{crit}}) \geq P_{\text{crit}} \\ \text{Definite-neutral} & \text{if } p_i(-T_{\text{crit}} < T_i < T_{\text{crit}}) \geq P_{\text{crit}} \\ \text{Ambiguous} & \text{if } \sigma_i \geq \sigma_\mu \\ \text{Residual-neutral} & \text{otherwise} \end{cases} . \quad (1)$$

We find it necessary to define additional ENSO states in order to account for the different reasons that lead to an inability to categorize an event as either El Niño (Figure 3a) or La Niña (Figure 3b). This can occur in several situations: either we are not sufficiently confident in the observation or the absolute temperature anomaly

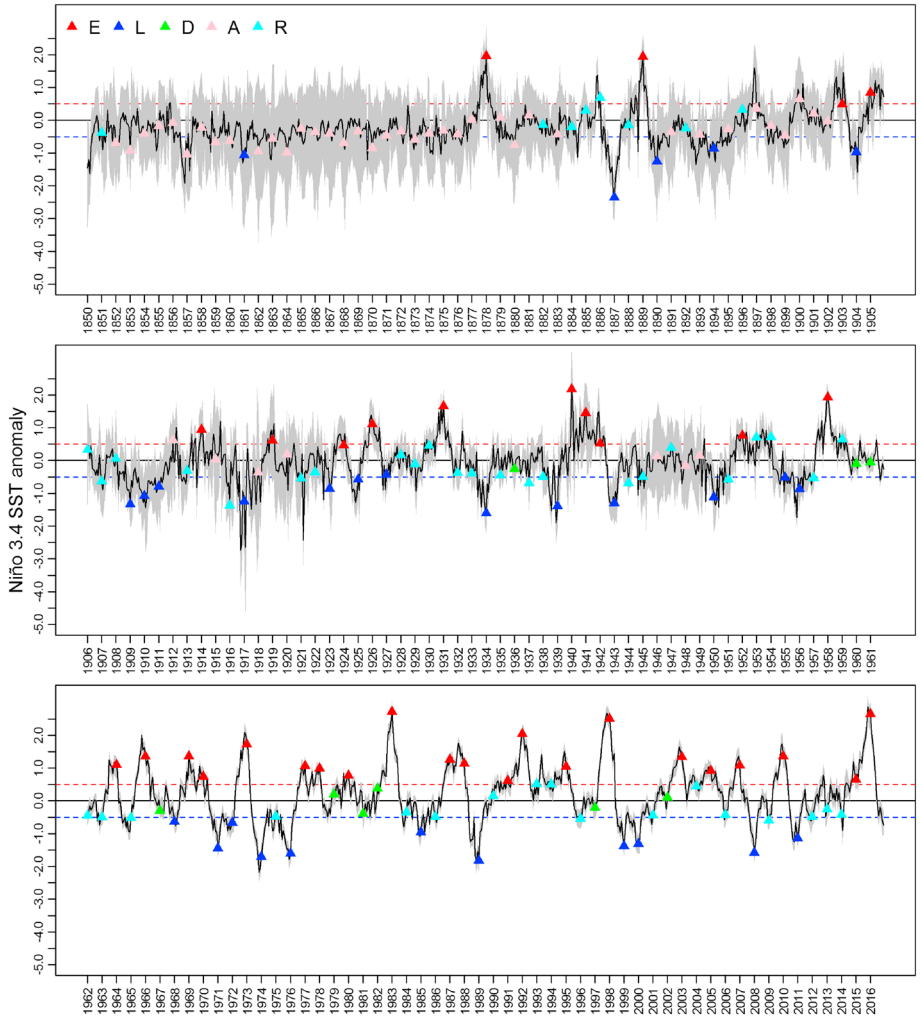

Figure 4. Time series of sea surface temperature anomalies in the Niño 3.4 region. The black line represents the median value, with the 95% confidence interval shown in gray. Blue and red dashed lines show the ±0.5°C critical values (equation (1). The classifications of the ENSO state during DJF for each year are shown by the triangles; with El Niño (E) in red, La Niña (L) in blue, definite-neutral (D) in green, ambiguous (A) in pink, and residual-neutral (R) in cyan.

is not sufficiently large. We consciously split our neutral state into a "definite" classification (where we have a strong knowledge that the temperature anomaly is less than T_{crit}, Figure 3c) and a "residual" classification (which often occurs during a transition into or out of an El Niño or La Niña, Figure 3e). We also introduce a new ENSO state that we term "ambiguous," to highlight years where one cannot feel confident enough to make a meaningful classification (Figure 3d). We consider that if the uncertainty in an individual temperature estimate is so large that it could simultaneously encompass both an El Niño and La Niña, any resulting ENSO state should be treated with skepticism. We express this in equation (1) as the standard deviation of a Niño 3.4 SST anomaly's PDF being greater than the standard deviation of index during the climatological period, σ_μ. In a climate modeling context, T_{crit} is often taken as $0.5\sigma_\mu$, which would collapse our definition down further [*Leloup et al.*, 2008]. It is worth stressing that the order in which these conditions are tested matters. For example, there is a large standard deviation in the estimated temperature anomaly for 1940 (Figure 4), but the warmth is so that a robust categorization can be made. In the results below, we present solely an annual classification of ENSO states derived from the December-January-February (DJF) season.

4.2. Past El Niño Years

The results of applying the probabilistic ENSO definition to the new ensemble are shown in Figure 4. It is apparent that we are rather certain of both the Niño 3.4 SST anomaly and the ENSO classification over the past 60 years. These classifications are consistent with other records derived from observations. It is also clear that

there is little knowledge of the ENSO state before 1880: nearly all years are classed as ambiguous (Figure 4). This accords well with *Hansen et al.* [2010] only starting their reconstruction from this date. Nonetheless, there are also several years since 1880 when the classification of ENSO is also ambiguous. Apart from a pause in the 1940s (e.g., 1946, Figure 3d) our knowledge of whether a year saw an El Niño or not appears robust beyond 1920. The different flavors of El Niño and their global impacts have become a burgeoning research topic of late [e.g., *Banholzer and Donner*, 2014]. The identification of a shift toward more central Pacific-type El Niños around the 1950s [*Yu and Kim*, 2013] seems potentially misplaced given the difficulty in detecting whether a year even had an El Niño (Figure 4).

5. Conclusion

Appropriate techniques to adequately infer regional temperatures from the limited early observations are essential to investigate the warming since the preindustrial. Here we applied a sophisticated approach to not only overcome the sparsity but to explicitly calculate the uncertainty in the resulting fields. The HadCRUT4 global temperature data set [*Morice et al.*, 2012] is used to demonstrate the utility of applying multiresolution lattice kriging [*Nychka et al.*, 2015]. Following the ethos of *Morice et al.* [2012], we have created a large ensemble of equally plausible spatially complete temperature fields to provide a quantified uncertainty for each location's past temperatures. The impact of these uncertainties was explored on a regional scale using temperatures in the Niño 3.4 region to describe past ENSO events. This necessitates a definition of El Niño that accounts for probabilistic temperature estimates. The uncertainty coming from a lack of nearby temperature observations can cause ambiguity in the classification of ENSO, especially prior to 1920.

Acknowledgments
This research was supported by a UCL Overseas Research Scholarship (M.I.), a University of the Punjab Overseas Scholarship (M.I.), and by NERC (NE/P006752/1) under the Belmont Forum's Pacmedy project (C.M.B). The helpful comments and suggestions by Doug Nychka are greatly appreciated. This research would not have been possible without the assistance and generosity of the Met. Office's Hadley Centre and the University of East Anglia's Climate Research Unit. The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work. The data generated and described in this research are freely available from the Oasis Hub (https://oasishub.co/).

References

Banholzer, S., and S. Donner (2014), The influence of different El Niño types on global average temperature, *Geophys. Res. Lett.*, *41*, 2093–2099, doi:10.1002/2014GL059520.
Berliner, M. L., D. Nychka, and T. Hoar (2012), *Studies in the Atmospheric Sciences*, vol. 144, Springer-Verlag, New York.
Brohan, P., J. J. Kennedy, I. Harris, S. F. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
Chang, K., S. Guillas, and V. Fioletov (2015), Spatial mapping of ground-based observations of total ozone, *Atmos. Meas. Tech.*, *8*, 3967–4009.
Cowtan, K., and R. G. Way (2014), Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, *Q. J. R. Meteorol. Soc.*, *140*(683), 1935–1944.
Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.
Hartmann, D., et al. (2013), *Observations: Atmosphere and Surface*, book section 2, pp. 159–254, Cambridge Univ. Press, Cambridge, U. K., and New York, doi:10.1017/CBO9781107415324.008.
Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto (2005), Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection, *Int. J. Climatol.*, *25*(7), 865–879.
Katzfuss, M. (2017), A multi-resolution approximation for massive spatial datasets, *J. Am. Stat. Assoc.*, *112*(517), 201–214.
Kleiber, W., and E. Porcu (2015), Nonstationary matrix covariances: Compact support, long range dependence and quasi-arithmetic constructions, *Stoch. Environ. Res. Risk Assess.*, *29*(1), 193–204.
Leloup, J., M. Lengaigne, and J.-P. Boulanger (2008), Twentieth century ENSO characteristics in the IPCC database, *Clim. Dyn.*, *30*(2), 277–291, doi:10.1007/s00382-007-0284-3.
Lindsey, R. (2013), *In Watching for El Niño and La Niña, NOAA Adapts to Global Warming*, National Oceanic and Atmospheric Administration, ClimateWatch Magazine, NOAA Climate.gov, United States.
Mastrandrea, M., et al. (2010), Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties. Intergovernmental Panel on Climate Change (IPCC). [Available at http://www.ipcc.ch.]
Morice, C., J. Kennedy, N. Rayner, and P. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.
Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain (2015), A multi-resolution Gaussian process model for the analysis of large spatial data sets, *J. Comput. Graphical Stat.*, *24*(2), 579–599.
Rohde, R., R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickhams, and S. Mosher (2013), Berkeley Earth temperature averaging process, *Geoinfor. Geostat. An Overview*, *1*(2), 1000103.
Ropelewski, C., and P. Jones (1987), An extension of the Tahiti–Darwin Southern Oscillation Index, *Mon. Weather Rev.*, *115*(9), 2161–2165.
Smith, T., R. Reynolds, T. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*(10), 2283–2296.
Trenberth, K. (1997), The definition of El Niño, *Bull. Am. Meteorol. Soc.*, *78*(12), 2771–2777.
Yu, J.-Y., and S. T. Kim (2013), Identifying the types of major El Niño events since 1870, *Int. J. Climatol.*, *33*(8), 2105–2112, doi:10.1002/joc.3575.