UCLA
UCLA Previously Published Works

Title
Discussion on “A combined estimate of global temperature”

Permalink
https://escholarship.org/uc/item/93d084b6

Journal
Environmetrics, 33(3)

ISSN
1180-4009

Author
McKinnon, Karen A

Publication Date
2022-05-01

DOI
10.1002/env.2721

Peer reviewed

eScholarship.org
Powered by the California Digital Library
University of California

Check for updates

Received: 22 December 2021 | Accepted: 18 February 2022

DOI: 10.1002/env.2721

DISCUSSION PAPER

WILEY

Discussion on "A combined estimate of global temperature"

Karen A. McKinnon

Department of Statistics, Institute of the Environment and Sustainability, and Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, California, USA

Correspondence
Karen A. McKinnon, Department of Statistics, Institute of the Environment and Sustainability, and Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095, USA.
Email: kmckinnon@ucla.edu

I thank the Editor for the invitation to comment on the recent paper by Craigmile and Guttorp (hereafter CG21) regarding a precision-weighted estimate of the global mean temperature anomaly time series.

Proper statistical modeling of global environmental data is challenging, due to both imperfect underlying data and the nonstationary covariance structure on a sphere. Weather station-based observations are nonrandomly distributed, both in terms of their surrounding environment (e.g., urban vs. rural) and their specific location in space, with large swaths of Africa and South America, as well as high-latitude locations, containing few if any weather stations that contribute to global land-based databases like the Global Historical Climatology Network (Menne et al., 2018). Similarly, ocean temperature data have historically been collected primarily along shipping routes, or were available at fixed moorings. Further, the temperature measurements themselves can contain artifacts unrelated to climate: weather stations can be moved from one location to another and/or the underlying land cover can change; temperature data are often recorded with different rounding conventions (Rhines et al., 2015); and different protocols across ships can lead to country-specific biases in sea surface temperature measurements (Chan et al., 2019). While many of these data issues have been corrected for monthly temperature data, others may remain. Finally, as noted above, the covariance structure of temperature is not expected to be stationary or isotropic across the globe, and performing inference on a global dataset with thousands of observations can become computationally challenging.

Historically, major governmental organizations have produced time series of global mean temperature anomalies using a wide range of statistical approaches, from simple bin averages (HadCRUT, although the latest version, HadCRUT5, uses geostatistical methods to extrapolate to data-sparse regions; Morice et al., 2021) to assuming a globally fixed "radius of influence" of 1200 km for each station, leading to substantial spatial smoothing (GISS; Hansen et al., 2010). More recently, Berkeley Earth, a nongovernmental organization, began producing its own estimates of global mean temperature anomalies using Gaussian process regression (Rohde & Hausfather, 2020), with a particular focus on whether issues with station data (e.g., urbanization around stations) were influencing trends in global mean temperature anomalies, which they did not find evidence for (Wickham et al., 2013). Perhaps surprisingly, given the wide range of statistical methods and, in some cases, use of slightly different underlying datasets, estimates of global mean temperature anomalies from different organizations are astoundingly similar, as is shown in figure 1 of CG21.

The contribution of CG21 is to take a more quantitative look at the shared structure and differences between five different commonly used global mean temperature anomaly datasets: Berkeley (Rohde & Hausfather, 2020), HadCRUT5 (Morice et al., 2021), NOAA (Vose et al., 2012), GISS (Hansen et al., 2010), and JMA (Ishihara, 2006). The authors propose a hierarchical Bayesian model that describes each global mean temperature anomaly time series as the sum of the desired latent global mean temperature anomaly time series (Y_t), and two noise terms ($\delta_{j,t} + \epsilon_{j,t}$), where j indexes the product and t indexes time, in years. The latent global mean temperature anomaly is itself a function of a slowly varying mean and a stationary Gaussian AR(4) process. The variance of the second noise term, $\epsilon_{j,t}$, is taken from the uncertainty measurement provided in each dataset, whereas $\delta_{j,t}$ (called the discrepancy term) is inferred in the modeling process.

What do we learn from the model? In my view, given the strongly shared signal between the raw value of each data product, the most interesting insights are in the error terms. It is encouraging to see the inferred positive correlations

 © 2022 John Wiley & Sons Ltd.

between the $\epsilon_{j,t}$ values across most of the products, excepting JMA, reflecting the shared sources of raw data. Of greater interest is the deviation term. Just as each data product uses quite different models to create their global mean temperature anomaly estimates, they also take different approaches in modeling the errors. As shown in figure 1b in CG21, three products (JMA, Berkeley, and GISS) have reasonably similar estimates of the uncertainty throughout their records, whereas NOAA provides a much larger estimate and HadCRUT5 suggests relatively high uncertainty earlier in the period with a large peak during World War II, but relatively small error by the end of the record. These differences could properly reflect different uncertainties for each product, and/or they could reflect different approaches to measuring uncertainty; the deviation term allows for accounting of differences within each data product that are not captured by these provided error estimates. While the 95% credible intervals of the discrepancy terms are small (generally $\pm 0.1°$C) compared to the global mean temperature anomaly, they can be comparable or larger than the standard error (as measured by 2σ for comparison), indicating the potential importance of accounting for additional uncertainties specific to each dataset. That said, a proper interpretation of the deviation term is challenging without additional information about its relationship to $\epsilon_{j,t}$. For example, the 95% credible interval for the NOAA dataset is notably larger than the others, which the authors interpret as indicating that the provided standard errors in NOAA are too large, or the others are too small. This comment suggests that the authors find that $\epsilon_{\text{NOAA},t}$ and $\delta_{\text{NOAA},t}$ are anticorrelated, in that the larger deviation term counteracts the variability from $\epsilon_{\text{NOAA},t}$, although this is not mentioned in the text. In order to better understand the full magnitude of the observational error, then, it would be helpful to present additional analyses of the summed $\delta_{j,t} + \epsilon_{j,t}$ term.

It is interesting to note that some of the largest posterior mean discrepancies are in the recent period when data coverage is reasonably good, with HadCRUT5 and GISS generally exhibiting positive discrepancies whereas those from JMA are likely to be negative. The underestimate of global mean temperature anomalies in JMA, at least as compared to the inferred latent time series and the other datasets, is almost certainly related to treatment of data in the Arctic. The Arctic, although small in area, has been warming at a rapid pace compared to the rest of the globe, so can have an outsize influence on estimates of global mean temperature anomalies. As first highlighted in Cowtan and Way (2014) with respect to HadCRUT4 (the prior version of the HadCRUT dataset), calculating a global mean with missing data in the Arctic will bias global mean temperature anomalies to be low, because the true mean value across the Arctic cannot be represented by the mean across the remainder of the globe. The Berkeley, HadCRUT5, NOAA, and GISS datasets all use various statistical methods to infer Arctic data, whereas JMA continues to leave out areas with few observations, as was done in HadCRUT4. The authors argue that "removing JMA from [the AR6 report] can not be justified on statistical grounds," but it does seem justifiable from the perspective that a known issue (exclusion of Arctic data) leads to a negative bias.

As a climate scientist with a foot in statistics, I will conclude with some thoughts on how to make analyses such as these more accessible to the climate science community, if desired. In general, it can be helpful if parameters inferred in the statistical modeling process can also have a valid scientific interpretation. For example, the authors divide the latent global mean temperature anomaly into a slowly varying mean term, μ_t, and an autocorrelated residual component, ν_t. This is potentially analogous to a common goal in climate science of dividing an observed time series or spatiotemporal field into a "climate change signal" and "internal variability noise," where the latter is driven by large-scale modes of climate variability, such as the El Niño-Southern Oscillation (Haustein et al., 2019). However, it is unlikely that the decomposition proposed by the authors also reaches this scientific goal. As noted in the final sentence of the paper, an appealing way to do so is to draw on information from climate models. In particular, single-model initial condition large ensembles (*Deser* et al., 2020a) and single-forcing simulations (e.g., Deser, Phillips, et al., 2020) can allow for clear separation of the climate change signal from internal variability, and could be used as additional information in developing the model for the latent global temperature anomaly. Relatedly, it should be noted that the term "natural variability," while used in CG21 to represent observational errors, almost always refers to the internal, dynamic variability within the climate system to climate scientists.

In addition, many climate scientists remain only minimally familiar with Bayesian methods, so it could be advantageous to "unpack" some of the methodological choices and results. In particular, describing the logic behind the choice of priors, and showing single samples from the MCMC sampler rather than only summary statistics would be helpful. The latter approach would help make the results more concrete for those unfamiliar with sampling methods, and also address prior concerns regarding missing information about the correlation structure between the two error terms.

In sum, the Bayesian hierarchical model for combining global mean temperature anomaly estimates from different research groups confirms the strength of the shared signal (the underlying latent global temperature time series), despite the use of a range of different methods across groups. Future work could clarify the behavior of the full error term, and make stronger links to scientific interpretations of each term in the model.

ORCID

Karen A. McKinnon https://orcid.org/0000-0003-3314-8442

REFERENCES

Chan, D., Kent, E. C., Berry, D. I., & Huybers, P. (2019). Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, *571*(7765), 393–397.

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., & Kay, J. E. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286.

Deser, C., Phillips, A. S., Simpson, I. R., Rosenbloom, N., Coleman, D., Lehner, F., Pendergrass, A. G., DiNezio, P., & Stevenson, S. (2020). Isolating the evolving contributions of anthropogenic aerosols and greenhouse gases: A new CESM1 large ensemble community resource. *Journal of climate*, *33*(18), 7835–7858.

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4), 1–29.

Haustein, K., Otto, F. E., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., Way, R. G., White, B., Subramanian, A., & Schurer, A. P. (2019). A limited role for unforced internal variability in twentieth-century warming. *Journal of Climate*, *32*(16), 4893–4917.

Ishihara, K. (2006). Calculation of global surface temperature anomalies with COBE-SST. *Weather Service Bulletin*, *73*, S19–S25.

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018). The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, *31*(24), 9835–9854.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P., & Simpson, I. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361.

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.

Rohde, R. A., & Hausfather, Z. (2020). The Berkeley earth land/ocean temperature record. *Earth System Science Data*, *12*(4), 3469–3479.

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J. H., Menne, M. J., Peterson, T. C., & Reynolds, R. W. (2012). NOAA's merged land–ocean surface temperature analysis. *Bulletin of the American Meteorological Society*, *93*(11), 1677–1685.

Wickham, C., Rohde, R., Muller, R. A., Wurtele, J., Curry, J., Groom, D., Jacobsen, R., Perimutter, S., Rosenfeld, A., & Mosher, S. (2013). Iinfluence of urban heating on the global temperature land average using rural sites identified from modis classifications. *Geoinfor Geostat: An Overview*, *1*(2), 1–6.

How to cite this article: McKinnon, K. A. (2022). Discussion on "A combined estimate of global temperature". *Environmetrics*, *33*(3), e2721. https://doi.org/10.1002/env.2721