# LMR Dataset

## Sources

[Published Paper](https://cp.copernicus.org/articles/15/1251/2019/cp-15-1251-2019.html)

While there is detailed discussion of proxy modeling and data processing, there is less emphasis placed on how the prior was formed. Considering how this paper is concerned with the second version of the LMR, it is full of references and comparison to the first version.

[Updates to LMR V2](https://datascience.codata.org/article/10.5334/dsj-2019-002/)

## Methodology

As stated in the paper, there are three major components used to generate the dataset. First is proxy records, which provide known data, climate model priors, to act as the estimated basis, and system models that connect and evaluate the relationship between the prior and proxy data.

This version of the dataset used an “offline” data assimilation approach, with prior ensemble data drawn from pre-existing model simulations. The simulations in question are the 2012 Coupled Model Intercomparison Project phase 5 (CMIP5) Last Millennium simulation by the Community Climate System Model version 4 (CCSM4) and the atmosphere-ocean-sea-ice model. The simulation here stretches from 850 to 1850 CE. The prior ensemble was constructed using random model states in the form of climate measurement deviations/anomalies sampled from the simulation, which was used for all years of the dataset reconstructions. This means that the prior did not contain any temporally specific climate data, disregarding specific climate events and trends. All trends and specific anomalies would need to be derived from the proxy data, with the prior providing no basis for them. Localization was applied to the proxy data, so that their influence on the data for a region decreased the further away they were.

In comparison with the LMR V1, 2250 additional proxy sites were added, but these are mostly concentrated in North America, Europe, the Arctic Circle, South America, and New Zealand. Only two proxies were added in the East Africa region, with a coral record in Madagascar and a tree ring record in the approximate location of Zimbabwe.

The author notes that averaging over Monte Carlo realizations rather than use any one specific one generally lead to more accurate results.

## Validation

The results of the LMR V2 dataset were evaluated against various other 20th century instrumental data records and reanalyses. The skill score metrics used were complementary, being the Pearson correlation coefficient and the coefficient of efficiency (CE). Spatial verification was performed by comparing the LMR temperature data using the [Berkeley Earth](https://www.scitechnol.com/berkeley-earth-temperature-averaging-process-IpUG.php?article_id=582) temperature analysis. The Palmer Drought Severity Index (PDSI) field was verified against the [2011 Dai](https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2010JD015541) data, which covers the years 1900-2008. I will note that according to this paper, the PDSI generally ranges between -10 (dry) to +10 (wet), with values below -3 indicating severe drought.

List of Sources Evaluated Against:

* NASA GISTEMP
  + Used For: Temperature (time?) Anomaly
  + Format: NCD
* MLOST
  + Used For: Air
  + Format: NCD
* NOAAGlobalTemp
  + Used For: Air
  + Format: NCD
* HadCRUT
  + Used For: Temperature Anomaly
  + Format: NCD
* BerkeleyEarth
  + Used For: Temperature
  + Format: NCD
* NOAA
  + Used For: Data
  + Format: NCD
* GPCC
  + Used For: Precipitation
  + Format: NCD
* DaiPDSI
  + Used For: PDSI
  + Format: NCD
* SPEI
  + Used For: SPEI
  + Format: NCD
* ICESM Last Milleium
  + Used For: tas\_sfc\_Amon, psl\_sfc\_Amon, pr\_sfc\_Amon, d18O\_sfc\_Amon, \_sfc\_Omon, sos\_sfc\_Omon, d18Osw\_sfc\_Omon
  + Format: NCD
* ICESM Last Millennium Historical
  + Used For: tas\_sfc\_Amon, psl\_sfc\_Amon, pr\_sfc\_Amon, d18O\_sfc\_Amon, \_sfc\_Omon, sos\_sfc\_Omon, d18Osw\_sfc\_Omon
  + Format: NCD
* CCSM4 Last Millenium
  + Used For: tas\_sfc\_Amon, psl\_sfc\_Amon, zg\_500hPa\_Amon, wap\_500hPa\_Amon, wap\_700hPa\_Amon, wap\_850hPa\_Amon, gmt\_sfc\_Amon, nhmt\_sfc\_Amon, shmt\_sfc\_Amon, pr\_sfc\_Amon, prw\_int\_Amon, scpdsi\_sfc\_Amon, scpdsipm\_sfc\_Amon, ai\_sfc\_Aann, pet\_sfc\_Amon, rlds\_sfc\_Amon, rlus\_sfc\_Amon, rsds\_sfc\_Amon, rsus\_sfc\_Amon, rlut\_toa\_Amon, rsdt\_toa\_Amon, rsut\_toa\_Amon, hfss\_sfc\_Amon, hfls\_sfc\_Amon, uas\_sfc\_Amon, vas\_sfc\_Amon, tos\_sfc\_Omon, sos\_sfc\_Omon, ohc\_0-700m\_Omon, ohcArctic\_0-700m\_Omon, ohcAtlanticNH\_0-700m\_Omon, ohcAtlanticSH\_0-700m\_Omon, ohcAtlantic\_0-700m\_Omon, ohcPacificNH\_0-700m\_Omon, ohcPacificSH\_0-700m\_Omon, ohcPacific\_0-700m\_Omon, ohcIndian\_0-700m\_Omon, ohcSouthern\_0-700m\_Omon, nheatGlobal\_Omon, nheatAtlanticArctic\_Omon, AMOCstreamfct\_Omon, AMOCindex\_Omon, AMOC26Nmax\_Omon, AMOC26N1000m\_Omon, AMOC45N1000m\_Omon
  + Format: NCD
* CCSM4 Preindustrial Control
  + Used For: AMOC26N1000m\_Omon, AMOC26Nmax\_Omon, AMOC45N1000m\_Omon, AMOCindex\_Omon, pr\_sfc\_Amon, psl\_sfc\_Amon, tas\_sfc\_Amon, zg\_500hPa\_Amon
  + Format: NCD
* CSSM4 Isotope Control Run
  + Used For: d180\_sfc\_Amon, tas\_sfc\_Amon, zg\_500hPa\_Amon
  + Format: NCD
* GFDL-CM3 Preindustrial Control
  + Used For: tas\_sfc\_Amon, zg\_500hPa\_Amon, AMOCstreamfct\_Omon,AMOCindex\_Omon, ohcAtlanticNH\_0-700m\_Omon, ohcAtlanticSH\_0-700m\_Omon, ohcAtlantic\_0-700m\_Omon, ohcIndian\_0-700m\_Omon, ohcPacificNH\_0-700m\_Omon, ohcPacificSH\_0-700m\_Omon, ohcPacific\_0-700m\_Omon, ohcSouthern\_0-700m\_Omon, ohc\_0-700m\_Omon, ohcArctic\_0-700m\_Omon, AMOC26N1000m\_Omon, AMOC45N1000m\_Omon, AMOC26Nmax\_Omon, psl\_sfc\_Amon, pr\_sfc\_Amon, wap\_1000hPa\_Amon, wap\_250hPa\_Amon, wap\_500hPa\_Amon, wap\_700hPa\_Amon, wap\_850hPa\_Amon, uas\_sfc\_Amon, vas\_sfc\_Amon, ua\_1000hPa\_Amon, \_250hPa\_Amon, ua\_500hPa\_Amon, ua\_700hPa\_Amon, ua\_850hPa\_Amon, ua\_925hPa\_Amon, va\_1000hPa\_Amon, va\_250hPa\_Amon, va\_500hPa\_Amon, va\_700hPa\_Amon, va\_850hPa\_Amon, va\_925hPa\_Amon
  + Format: NCD
* MPI-ESM-P Last Millennium
  + Used For: msftmyz\_Omon, tas\_sfc\_Amon, zg\_1000hPa\_Amon, zg\_500hPa\_Amon, AMOCindex\_Omon, AMOC26N1000m\_Omon, AMOC45N1000m\_Omon, AMOC26Nmax\_Omon, AMOCstreamfct\_Omon, psl\_sfc\_Amon, pr\_sfc\_Amon, uas\_sfc\_Amon, vas\_sfc\_Amon, wap\_500hPa\_Amon, wap\_700hPa\_Amon, wap\_850hPa\_Amon, wap\_250hPa\_Amon, wap\_1000hPa\_Amon, ua\_1000hPa\_Amon, ua\_250hPa\_Amon, ua\_500hPa\_Amon, ua\_700hPa\_Amon, ua\_850hPa\_Amon, \_925hPa\_Amon, va\_1000hPa\_Amon, va\_250hPa\_Amon, \_500hPa\_Amon, va\_700hPa\_Amon, va\_850hPa\_Amon, va\_925hPa\_Amon
  + Format: NCD
* 20 CR
  + Used For: tas\_sfc\_Amon, zg\_1000hPa\_Amon, zg\_500hPa\_Amon, psl\_sfc\_Amon, wap\_1000hPa\_Amon, wap\_250hPa\_Amon, wap\_500hPa\_Amon, wap\_700hPa\_Amon, wap\_850hPa\_Amon, pr\_sfc\_Amon, prw\_int\_Amon, uas\_sfc\_Amon, vas\_sfc\_Amon, ua\_1000hPa\_Amon, ua\_250hPa\_Amon, ua\_500hPa\_Amon, \_700hPa\_Amon, ua\_850hPa\_Amon, ua\_950hPa\_Amon, va\_1000hPa\_Amon, va\_250hPa\_Amon, va\_500hPa\_Amon, va\_700hPa\_Amon, va\_850hPa\_Amon, va\_950hPa\_Amon, ua\_600hPa\_Amon, va\_600hPa\_Amon
  + Format: NCD
* Era 20C
  + Used For: psl\_sfc\_Amon, tas\_sfc\_Amon, pr\_sfc\_Amon, prw\_int\_Amin, uas\_sfc\_Amon, vas\_sfc\_Amon, wap\_1000hPa\_Amon, wap\_250hPa\_Amon, wap\_500hPa\_Amon, wap\_700hPa\_Amon, wap\_850hPa\_Amon, zg\_500hPa\_Amon, ua\_1000hPa\_Amon, ua\_250hPa\_Amon, ua\_500hPa\_Amon, ua\_700hPa\_Amon, ua\_850hPa\_Amon, ua\_925hPa\_Amon, va\_1000hPa\_Amon, va\_250hPa\_Amon, va\_500hPa\_Amon, va\_700hPa\_Amon, va\_850hPa\_Amon, va\_925hPa\_Amon, ua\_600hPa\_Amon, va\_600hPa\_Amon, ua\_950hPa\_Amon, va\_950hPa\_Amon
  + Format: NCD
* Era 20cm
  + Used For: tas\_sfc\_Amon, zg\_500hPa\_Amon
  + Format: NCD
* Loveclim Goosse 2005
  + Used For: tas\_sfc\_Amon, tos\_sfc\_Amon
  + Format: NCD
* IHADcm3 Preindustrial Control
  + Used For: tas\_sfc\_Amon, psl\_sfc\_Amon, pr\_sfc\_Amon, d180\_sfc\_Amon
  + Format: NCD
* CCSM3 Trace 21KA
  + Used For: tas\_sfc\_Amon, psl\_sfc\_Adec, sfcheight\_sfc\_Adec, tas\_sfc\_Adec, ts\_sfc\_Adec, tsl\_sfc\_Adec, prw\_int\_Adec, pr\_sfc\_Adec, zg\_500hPa\_Adec, zg\_850hPa\_Adec, wap\_700hPa\_Adec, ta\_1000hPa\_Adec, ta\_850hPa\_Adec, ta\_700hPa\_Adec, ta\_600hPa\_Adec, ta\_500hPa\_Adec, tos\_sfc\_Odec, sos\_sfc\_Odec
  + Format: NCD
* CGenie PETM
  + Used For: tas\_sfc\_Adec, tos\_sfc\_Adec
  + Format: NCD

Non-Default Datasets

* Old 20CR
  + Used For: Air
  + Format: NCD
  + Purpose: LIM Calibrations in Perkins and Hakin 2017
* Shortened Berkeley Earth
  + Used For: tas\_run\_mean
  + Format: NCD
  + Purpose: Shortened dataset for more observational coverage
* Test 20CR
  + Use For: Air, time series (?)
  + Format: NCD
  + Purpose: Test data for test gridded variable

The spatially-indexed PDSI field has an globally-averaged correlation coefficient of 0.09 and CE of 0.00. These are much lower than the skill scores of the other fields provided, such as spatial temperature, which had a correlation coefficient of 0.52 and CE of 0.22. In comparison to LMR V1, the additional proxy data mostly affected the United States and western Europe, and most of the impact came from tree-ring width records, not coral and ice core records. Considering how, as previously stated, most of the additional tree-ring records are concentrated in North America, Western Europe, the Arctic circle, and North America, this may be cause for concern that not only does the region around East Africa and much of the Indian Ocean world have les proxy data, but that the proxy data that does exist is not as useful. However, as seen in Figure 10, the proxy data does seem to have positively affected the correlation coefficient in East Africa at the very least, though the CE score is mostly unchanged.

For temporally-indexed data, across all verification metrics, there was a general increase in skill score over time. A specific example would be another verification metric used – the difference in CE score between the uninformed prior and posterior/reanalysis dataset. For the years covering 1500-1879, the overall difference in CE was 0.10, and was 0.13 in 2000. This upwards trend indicates a larger usage of proxy records over the prior in the latter years of the Common Era. For years 1-499 CE, there was a practically negligible difference of 0.00, indicating that the prior and posterior models were more or less the same, and that there was little proxy data build with.

Despite overall improvements in correlation and CE scores in comparison to LMR V1, there is some decrease in skill in parts of the southern Atlantic and Indian oceans, though it was considered acceptable thanks to the much larger improvements elsewhere.

In regards to the skill score collections displayed in the paper, the paper mostly focuses on temperature and Berkeley Earth for verification. Verification for PDSI was mentioned, but was shown in less detail, and no spatial verification figures was directly provided. They make little mention at all of precipitation, and do not discuss how this field was verified. This paper is also not as useful to evaluate the LMR V2 dataset itself, as most of it is focused on improvements with regard to LMR V1. Considering this, I have decided to review the LMR V1 publication as well, to see the initial starting points and gain a better understanding of the comparisons.

In the conclusions, the author suggests possible improvements to the PDSI through screening tree-ring proxies. In the current LMR V2 dataset, some skill improvement in the eastern United States has come with decreased accuracy for some important features of pre-industrial climate. They believe that with an updated methodology, the improvements can be kept without coming at the expense of the data in other regions or time periods.

### PDSI

In regards to the PDSI, the verification covered the approximate span of time of the Dai dataset, spanning from 1800-2000. The self-calibrated PDSI values were used, of which there were two types. These were the Thorntwaite and Penman-Monteith PDSI variables – representing the PDSI values from two different methodologies for deriving precipitation evapotranspiration (PET), an important variable for calculating PDSI. It should be noted that this should not affect the scale of the PDSI derived. According to <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2010JD015001>, the self-calibrating PDSI is viewed as generally more reliable for a global data set, considering how the unmodified Palmer index uses calibrating parameters derived from the United States. The self-calibrating PDSI index on the other hand, uses parameters specific to a particular location.

The mean of all Monte Carlo reconstructions was used to compare against the Dai dataset. Examining the code used to process the reconstruction and Dai data, it would appear that none are scaled or transformed in such a way that would affect the values themselves.

## Advantages

* **Everything Publicly Available:** As the goal of the project was to have a completely open-access and publicly available resource, all code is on GitHub and only publicly available or archived proxy data and other reconstructions are used.
  + **Code Available for Basic Reconstructions:** There is code and data available online for people to create their own basic reconstructions, without the need of heavy processing.
* **Good Documentation:** Several papers, websites, and thorough commenting in the code available for reference.

## Disadvantages

* **Seasonal Data Not Always Available:** Not all proxies have been screened and vetted in regard to their seasonality.
* **Only Annual Data Available:** Generally, only annual data is available in this dataset, making it more difficult to determine climate anomalies on a scale less than 12 months.
* **Low Resolution:** All data is spatially indexed on a 2 x 2 degree coordinate grid.