



# A new global reconstruction of temperature changes at the Last Glacial Maximum

J. D. Annan and J. C. Hargreaves

Research Institute for Global Change, Yokohama, Japan

Correspondence to: J. D. Annan (jdannan@jamstec.go.jp)

Received: 11 September 2012 – Published in Clim. Past Discuss.: 10 October 2012

Revised: 23 January 2013 – Accepted: 23 January 2013 – Published: 13 February 2013

**Abstract.** Some recent compilations of proxy data both on land and ocean (MARGO Project Members, 2009; Bartlein et al., 2011; Shakun et al., 2012), have provided a new opportunity for an improved assessment of the overall climatic state of the Last Glacial Maximum. In this paper, we combine these proxy data with the ensemble of structurally diverse state of the art climate models which participated in the PMIP2 project (Braconnot et al., 2007) to generate a spatially complete reconstruction of surface air (and sea surface) temperatures. We test a variety of approaches, and show that multiple linear regression performs well for this application. Our reconstruction is significantly different to and more accurate than previous approaches and we obtain an estimated global mean cooling of  $4.0 \pm 0.8$  °C (95 % CI).

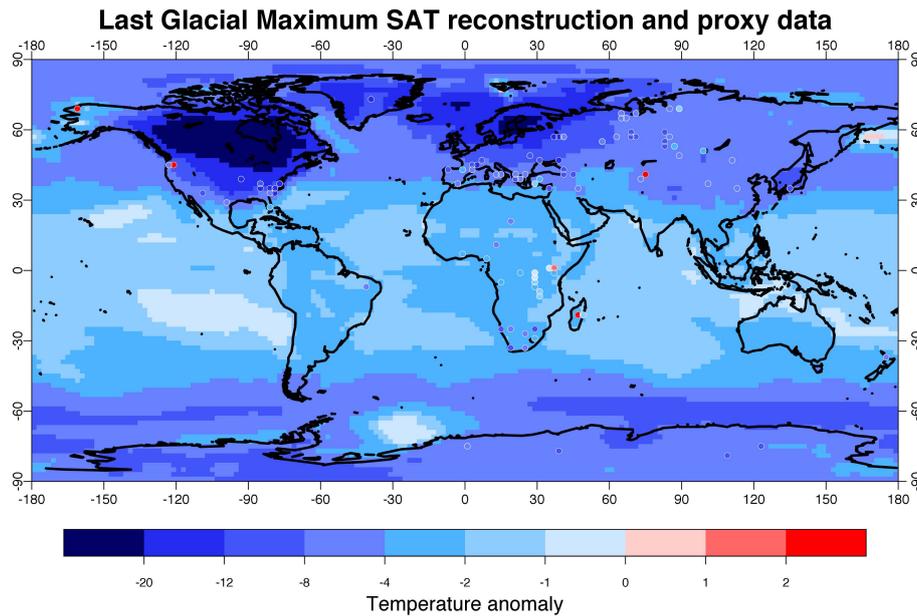
## 1 Introduction

The Last Glacial Maximum (LGM, 19–23 ka BP) represents the most recent interval when the global climate was substantially different to the present, and therefore provides us with a key target in testing the response of climate models to large changes in radiative forcing. There is, however, significant disagreement even over first-order diagnostics such as the global average of the annual mean temperature at that time, with estimates ranging from as much as 6 °C to as little as 3 °C colder than the modern (pre-industrial) climate (e.g. Schneider von Deimling et al., 2006a; Holden et al., 2009; Schmittner et al., 2011). This uncertainty limits our ability to critically assess climate model performance.

An early reconstruction of the global sea surface temperature (SST) anomaly at the LGM was made by the CLIMAP project (Climap Project Members, 1976), which estimated a

globally-averaged value of 2.3 and 0.8 °C in the tropics (all temperature anomalies are presented here as pre-industrial climate minus LGM). However, these values were argued to be substantially too small by subsequent analyses, which presented tropical LGM SST estimates of around 2.5–3 °C colder than present (Crowley, 2000; Ballantyne et al., 2005). Simulations of the LGM using state of the art atmosphere–ocean global climate models (GCMs) generally generate global mean surface air temperature (SAT) anomalies in the range of 3–5 °C colder than present (Braconnot et al., 2007), but these values are thought to be biased warm due to the experimental design, which omits the likely negative forcings of vegetation and dust changes (Crucifix and Hewitt, 2005; Schneider von Deimling et al., 2006a). These results can be interpreted as implying a model-based range of around 4–7 °C if these extra forcings were to be accounted for (Jansen et al., 2007).

The first attempts at directly constraining model results with proxy data produced results consistent with this range, with resulting best estimates for the global mean SAT anomaly of around 6 °C (Schneider von Deimling et al., 2006a; Holden et al., 2009). However, a new analysis has recently challenged this emerging consensus with a remarkably mild estimate of 3.0 °C (90 % range 1.7–3.7 °C) (Schmittner et al., 2011), based on the fit of an intermediate complexity climate model to the most recent comprehensive proxy syntheses. Such a mild climate state, if confirmed in other studies, would be difficult to reconcile with GCM simulations. The response of the climate system to a large forcing is of fundamental importance to understanding future climate change, and therefore the large discrepancy between these analyses requires further investigation.



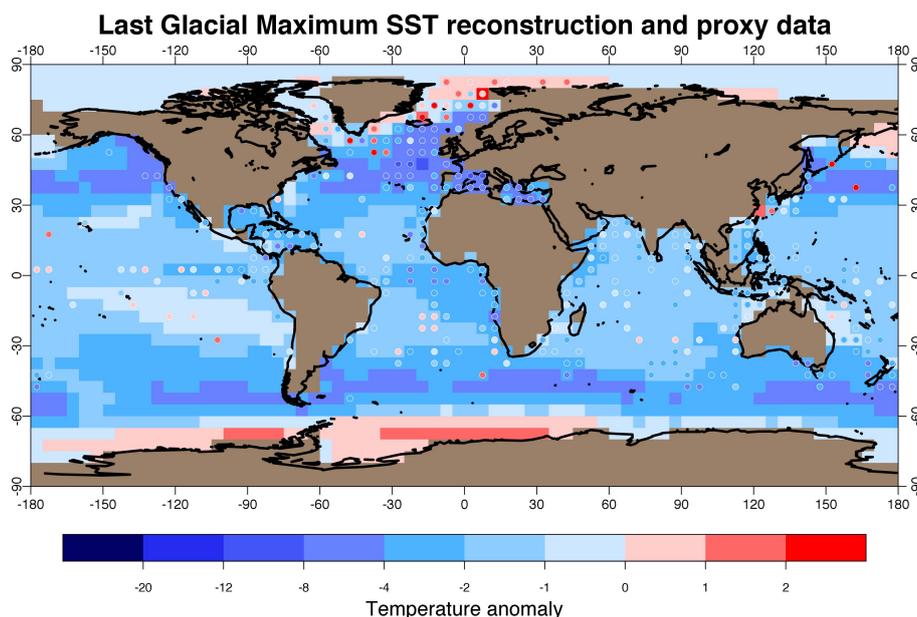
**Fig. 1.** Reconstruction of Last Glacial Maximum surface air temperature anomaly ( $^{\circ}\text{C}$ ) based on multi-model regression. Proxy data are represented as coloured dots.

In this paper we present a new model-data synthesis, combining the recent comprehensive compilation of proxy data as used by Schmittner et al. (2011), together with the ensemble of state of the art GCMs which participated in the PMIP2 project (Braconnot et al., 2007). The data and models are introduced more fully in Sect. 2. In Sect. 3 we describe several approaches to reconstructing the climate state: increasing complexity and accuracy. In order to test the reconstruction methods, and to estimate their uncertainties, we perform extensive cross-validation using each of the PMIP2 simulations in turn as the target, extracting pseudoproxy data from the appropriate locations, and calculating the accuracy of the resulting reconstruction based on these data. Confidence intervals are presented at the 95 % level based on the spread of cross-validation results unless otherwise stated. We start with an attempt to simply smooth the data in Sect. 3.1; this being a commonly used approach to generate climate field reconstructions. However, the sparseness of the data, and in particular the non-random nature of large data void areas, limits the performance of this approach. In Sect. 3.2 we consider the pattern scaling approach, in which a single model anomaly field is scaled to optimally fit to the data. While this method improves on the smoothing, the results are still rather moderate. Our main result, presented in Sect. 3.3, is based on multiple linear regression of the ensemble of climate model fields. This method performs substantially better than the other two approaches. Detailed validation and some sensitivity analyses are presented in Sect. 4. We summarise and discuss some implications of our result in Sect. 5.

## 2 Data and models

The proxy data which we use here consist of a multiproxy analysis of SST anomalies presented on a  $5^{\circ}$  grid (MARGO Project Members, 2009), and SAT anomalies on a  $2^{\circ}$  grid over land based on pollen and plant macrofossils (Bartlein et al., 2011), with some additional points from a variety of sources including Antarctic and Greenland ice cores (Shakun et al., 2012). The data are displayed as the dots in Figs. 1 and 2. While the land data of Bartlein et al. (2011) are provided with uncertainty estimates, the ocean data are not, instead being associated with a nondimensional “reliability index”. One common interpretation of this parameter is to treat it as the one standard deviation uncertainty of a Gaussian error (Hargreaves et al., 2011; Schmittner et al., 2011). Analyses presented in Sect. 4.2 cast some doubt on the accuracy of these uncertainty estimates, but our results are not sensitive to this factor.

We use the outputs of nine models which participated in the PMIP2 project (Braconnot et al., 2007), being all of those for which both atmosphere and sea surface temperatures are available. The models used for this were predominantly state of the art atmosphere–ocean GCMs, with some models also including an interactive vegetation component and one being an intermediate-complexity model with simplified atmosphere. Model outputs were typically calculated as 100 yr averages to minimise the effect of internal variability. The experimental protocol for the LGM accounts for the largest and best-quantified forcings at that time, which include reduced greenhouse gas concentrations, minor changes in orbital parameters, and extensive increases in Northern Hemisphere



**Fig. 2.** Reconstruction of Last Glacial Maximum sea surface temperature anomaly ( $^{\circ}\text{C}$ ) based on multi-model regression. Proxy data are represented as coloured dots. Land areas are masked as brown.

ice sheets. The experimental design and main results are described more fully by Braconnot et al. (2007). Despite some limitations in the forcing protocol (Schneider von Deimling et al., 2006a), the model outputs appear to generally provide a reasonable representation of the Last Glacial Maximum (Hargreaves et al., 2011). The global surface air temperature anomalies simulated by these models at the LGM range from 3.1 to 5.9  $^{\circ}\text{C}$  colder than present. Inter-model differences are particularly large over the ice sheets, to which a number of factors may contribute (Abe-Ouchi et al., 2007). All model data were regridded onto regular 2 and 5  $^{\circ}$  grids for SAT and SST, respectively, to match the proxy syntheses. In order to effectively combine the data with the the ensemble of models, we eliminated a small number of data points where, due to grid inconsistencies, either one or more models provided no SST output at the location of an SST data point, or where a pollen-derived SAT estimate was located under one or more of the models' ice sheets, leaving us a total of 309 SST points and 95 SAT points. Including the ice-covered points in our analysis only changes our result by less than 0.1  $^{\circ}\text{C}$  in the global mean.

### 3 Results

#### 3.1 Smoothing

The data set (which is presented in Figs. 1 and 2) gives widespread coverage such that 95 % of the surface of the Earth is within 2000 km of a data point. This good coverage might suggest that a direct smoothing – such as was

performed for the ocean alone by CLIMAP (Climap Project Members, 1976) and which is commonly used for modern temperature anomaly fields (Hansen and Lebedeff, 1987; Smith et al., 2008) – could give good results. However, we find this not to be the case. The performance of smoothing was investigated through the use of pseudoproxy data taken from the PMIP2 models. We tested distance-based weighted averaging to smooth the data over the full global grid, using both Gaussian and exponential weighting functions over a wide range of length scales. Best results were obtained when we smoothed over land and ocean data separately so as to maintain the land–ocean contrast, but ignoring SST data north of 50  $^{\circ}\text{N}$  and treating this region as land, due to the presence of sea ice which insulates the ocean from the overlying atmosphere (Hargreaves et al., 2011), using a Gaussian weighting with a length scale of 500 km. However, the results were rather insensitive to these choices. Applying this smoothing process to pseudoproxy data from the PMIP2 models generates rather mediocre results with an area-weighted pointwise RMS error over the globe of over 3.6  $^{\circ}\text{C}$ , and a bias in the global mean of the temperature field of  $-0.9 \pm 0.9$   $^{\circ}\text{C}$ . That is, the smoothing tends to strongly underestimate the overall cooling at the LGM. This is primarily due to the absence of observed data over the areas with the largest anomalies (particularly, where the massive Laurentide and Fennoscandian ice sheets covered the glacial Earth) which results in an extremely large underestimate of the cooling simulated in these areas.

When we smooth the proxy data in the same manner, the resulting field has a global mean SAT anomaly (under the common assumption that the SAT anomaly over the ocean

is equal to the SST anomaly) of  $3.2^{\circ}\text{C}$ . After correcting for the likely bias according to the pseudoproxy experiments, we therefore obtain an estimate of  $4.1 \pm 0.9^{\circ}\text{C}$  (see Table 1). This is consistent with our main result of Sect. 3.3, but the spatial pattern is very noisy and unrealistic. A notable advantage of this method, however, is that it makes minimal use of model output, only relying on it for an estimate of the bias due to spatially inadequate sampling. Therefore, we consider this calculation a useful confirmation of our main result. Shakun et al. (2012) obtained an estimate of  $3.6^{\circ}\text{C}$  based on interpolating a small subset of this data set, but made no attempt to estimate or correct for the bias due to sampling location.

Although more sophisticated smoothing methods such as kriging could in principle be applied to this problem, it is the large voids in the spatial distribution of data which lead to the large bias and mediocre performance, and we thus conclude that the spatial complexity of the cooling pattern requires a climate model to realistically represent it.

### 3.2 Pattern scaling through linear regression

Several researchers have addressed the question of the global temperature change at the LGM by fitting a climate model to proxy data (Schneider von Deimling et al., 2006a; Holden et al., 2009; Schmittner et al., 2011). Such an approach uses the model to extrapolate into data voids, thus ensuring physically plausible results across the globe. The model fitting is primarily performed by tuning internal model parameters which relate to the radiative feedback (and thus climate sensitivity) of the models. However, due to the computational cost of this approach, it can generally only be applied to models of intermediate complexity or resolution. We cannot directly simulate this approach with the PMIP2 models, as we only have the results of one simulation for each model, and cannot re-run them at multiple parameter settings. However, we can approximate the effect of changing their sensitivities by the pattern scaling approach (Santer et al., 1990) in which a linear scaling factor (estimated through linear regression with intercept fixed at zero) is applied to the anomaly fields. While pattern scaling is not as powerful and flexible as running an ensemble of simulations with multiple adjustable parameters, it should capture a dominant fraction of the response to changing the sensitivity of the model. The predicted climate anomaly field  $S$  is thereby estimated as

$$S = \alpha F \quad (1)$$

where  $F$  is the anomaly field generated by the model and  $\alpha$  is a scalar chosen so as to minimise the unweighted sum of squared residuals  $\sum_i (s_i - o_i)^2$  at the  $i$  locations where proxy-based estimates  $o_i$  exist. We do not use an area weighting for the fit. While the land data are presented on a  $2^{\circ}$  grid and the ocean on  $5^{\circ}$ , the number of cores which contribute to each non-empty grid box is roughly the same for each data

**Table 1.** Summary of global temperature anomaly estimates (all values in  $^{\circ}\text{C}$  and quoted as pre-industrial minus LGM).

Method	Raw value	Bias	Result
Smoothing	3.2	$-0.9 \pm 0.9$	$4.1 \pm 0.9$
Single model	3.9	$-0.6 \pm 1.5$	$4.5 \pm 1.5$
Multimodel	3.9	$-0.1 \pm 0.8$	$4.0 \pm 0.8$

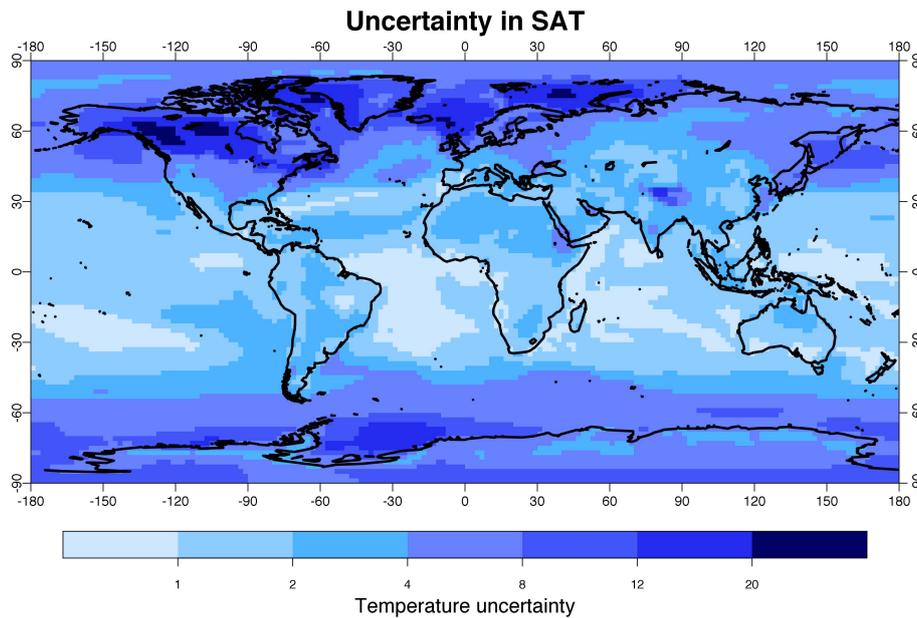
set, at 2–3 per grid box, with no strong latitudinal pattern. Therefore, we do not consider it appropriate to assign a much higher weight either to ocean versus land data, or low versus high latitude cells, as area weighting would imply.

One possible improvement to this methodology would be to explicitly account for observational uncertainty in the weighting. However, the sensitivity analyses discussed in Sect. 4.2 suggest that the uncertainty estimates may not be reliable. Moreover, this actually has negligible influence on our results.

The uncertainty of the pattern-scaling reconstruction is again estimated by cross-validation, using all model pairs for target and predictor. The area-weighted pointwise RMS error in temperature anomaly generated by this pattern scaling approach is lower than that of smoothing, at  $2.9^{\circ}\text{C}$ . The resulting error on the global average of the temperature anomaly is  $-0.6 \pm 1.5^{\circ}\text{C}$ , with the mean bias arising from the well-known phenomenon of regression attenuation or dilution (Snedecor and Cochran, 1989, Sect. 9.14). The underlying reason for this is that the target values were not actually generated by the addition of independent identically distributed errors to the predictor variables but rather both are approximations to the true climate, and thus the residuals tend to be smallest for the largest predicted anomalies and vice versa. While there are more sophisticated approaches to regression that can in principle account for this effect, they require additional assumptions, and here we prefer instead to use the estimate of attenuation bias obtained through cross-validation to correct our result accordingly. When we fit each model to the proxy data in turn and take the ensemble average, the resulting global mean temperature anomaly after bias correction is  $4.5 \pm 1.5^{\circ}\text{C}$ . Although the bias is reduced by this approach compared to smoothing, there is still substantial error both at the gridpoint level and in the global average of the temperature anomaly, which is due to the substantially different spatial anomaly patterns simulated by different climate models under glacial conditions. Thus, it appears that this method, while clearly superior to a simple smoothing, also has significant limitations.

### 3.3 Multiple linear regression

A natural extension of the previous method to the case of multiple heterogeneous models, is to use multiple linear regression, which has also been termed the “superensemble” (Krishnamurti et al., 2000):



**Fig. 3.** Uncertainty in Last Glacial Maximum surface air temperature anomaly ( $^{\circ}\text{C}$ ) from bootstrap resampling. Results presented as half-width of 95 % confidence interval.

$$S = \sum_j \alpha_j F_j \quad (2)$$

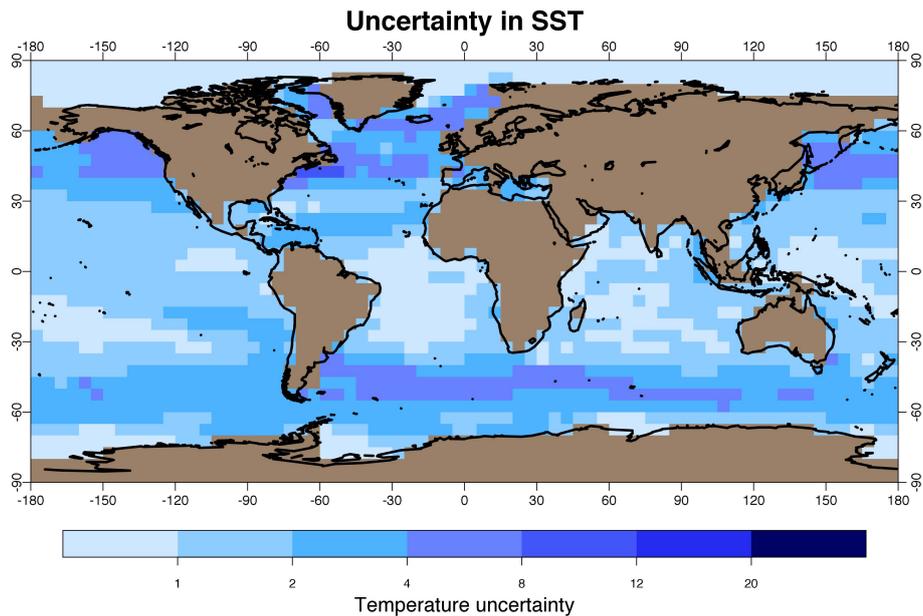
where the sum is over multiple models  $F_j$  and the scaling factors  $\alpha_j$  are chosen to minimise the sum of squared residuals as before.

One important feature of this method, as opposed to probabilistic weightings such as Bayesian Model Averaging (Hoeting et al., 1999), is that the scaling factors applied to the models here are not constrained either to be positive or even to sum to unity. The method does not treat the models as prior estimates of the climate state but merely as a set of possible predictors for it, and the result is not constrained to lie within the ensemble range, but instead it can be any arbitrary linear combination of the different spatial patterns that the individual models exhibit. Again, due to our concerns about the estimated proxy errors, we implement a simple unweighted regression and use cross-validation within the PMIP2 ensemble to estimate the uncertainties. That is, each model in turn was selected as the target, pseudoproxy data simulated from it, and the remaining models used as the predictor set. As a confirmation of algorithmic correctness, we also checked that including the target model in the predictor set invariably results in a near-perfect reconstruction, even when imperfect observations, are used.

Our main results are presented in Figs. 1 and 2, and the reconstructed fields are also available as Supplement. Their pointwise uncertainties are shown in Figs. 3 and 4. In contrast to smoothing or single model scaling, multiple linear regression generates a very small bias of  $-0.1 \pm 0.8^{\circ}\text{C}$  in the global mean, and the area-weighted pointwise RMS error is also much lower, at  $2.0^{\circ}\text{C}$ .

Our estimated global average of the annual mean surface air temperature anomaly is  $4.0 \pm 0.8^{\circ}\text{C}$ , and is summarised in Table 1 along with the results of the two other methods tested here. As expected, SST and SAT anomalies show good agreement over open water, but the two fields diverge strongly at high latitudes due to sea ice, with the reconstructed SST field showing slight warming both at high northern latitudes (in agreement with proxy data) and around Antarctica (where there are no observations). Fields of uncertainties are shown in Figs. 3 and 4. The marginal warming in the Barents sea area in the SAT reconstruction is in a region of very high uncertainty, as there are few proxy data for SAT close to this region, and model simulations disagree substantially here. Thus, we have low confidence that this is a genuine feature of the climate system. However, over most of the globe, the estimated cooling is substantially greater than its associated uncertainty.

Uncertainty in the reconstruction is particularly low across the tropical region, but increases significantly with latitude. There is negligible latitudinal trend in the residuals over either ocean or land, but they are generally positive over land (average  $0.6^{\circ}\text{C}$ ) and negative over the ocean (average  $-0.2^{\circ}\text{C}$ ), suggesting a tension between the land and ocean data which the models struggle to represent. We therefore tested the robustness of our result by considering only ocean or land data in turn. These two data sets result in estimates of  $3.5 \pm 1.2^{\circ}\text{C}$  and  $4.6 \pm 0.8^{\circ}\text{C}$ , respectively, which although not in close agreement, are consistent with each other, and our main result, within their respective uncertainties. Each of these two results, however, would imply a substantial mean bias in the withheld data set, of  $0.8^{\circ}\text{C}$  in SAT data (when



**Fig. 4.** Uncertainty in Last Glacial Maximum sea surface temperature anomaly ( $^{\circ}\text{C}$ ) from bootstrap resampling. Results presented as half-width of 95 % confidence interval. Land areas are masked as brown.

only SST data were used) and  $-1.1^{\circ}\text{C}$  in SST data (when only SAT were used). Biases of this magnitude seem unlikely given the comprehensive multiproxy consensus that each of these data sets represents. The land–sea contrast in the data (a ratio of 3.2 between the respective means) is marginally larger than that found in any of the PMIP2 models (which range from 1.9 to 3.1), and perhaps more plausible explanations for this are that inadequacies in forcings (such as atmospheric dust, or vegetation feedbacks), or model physics, might cause an underestimate of the land/ocean amplification in the model simulations. As a sensitivity test, we repeated the calculations after reducing all modelled temperatures over land uniformly by  $1^{\circ}\text{C}$ . This increases the land–ocean contrasts of the models to a level more comparable to that of the data, and the results obtained when using only ocean data ( $3.8^{\circ}\text{C}$ ), or land data ( $4.1^{\circ}\text{C}$ ), are in much closer agreement with each other.

Further tests using smaller ensembles are described more fully in Sect. 4.1, and provide no evidence of either overfitting or inadequacy in the ability of multiple linear regression to adequately describe the global climate system. While we cannot rule out the possibility that future model development (including experimental design such as a more complete set of forcings) could lead to a slightly different result, our sensitivity tests suggest that our result is largely insensitive to modelling uncertainties. On the other hand, any major re-evaluation of the proxy data (such as has happened in the past for tropical temperatures) could potentially affect our result, but conditional on the data analysis, our result appears to be highly robust.

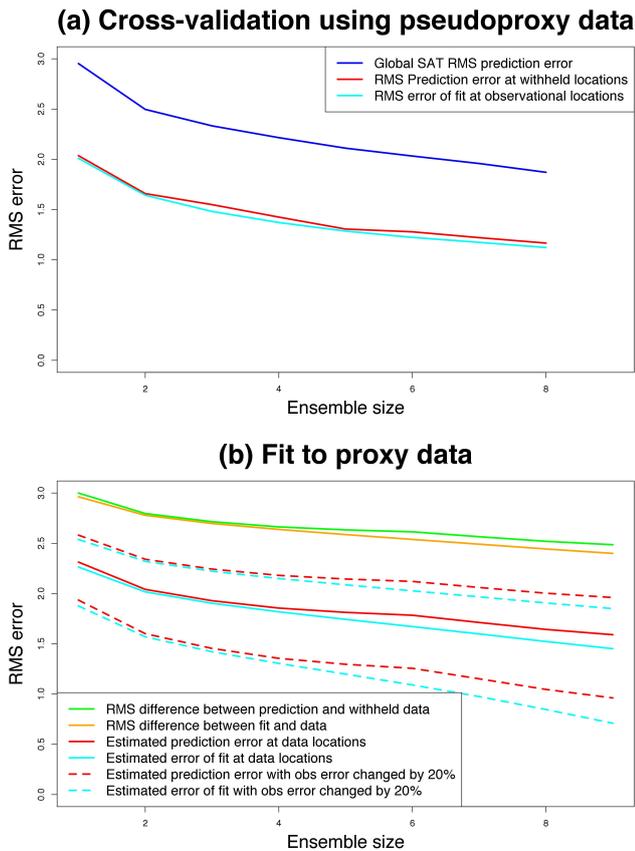
## 4 Sensitivity analyses and validation

### 4.1 Validation of multiple linear regression

The basis of our method is the use of the ensemble of models (each of which is expected to provide a physically plausible depiction of the climate system’s response to LGM forcing) as a set of possible predictors, with linear regression used to find the optimal combination of these predictors.

One obvious problem that could arise with this method is that of overfitting. In-sample performance can only improve with additional predictors even if they are nonsensical or random, but this may not lead to an improved global reconstruction. Thus, we use pseudoproxy experiments to investigate how the results vary with ensemble size. We randomly select one model as the target, and use random subsets of the ensemble in the multiple linear regression. Pseudoproxy data are sampled from the output fields of the target model at the same locations as the real data. Additionally, one of these data points (again randomly selected) is withheld from the regression in order that we can check the predictive performance at the data locations, as distinct from over the entire globe. Overfitting, were it to occur with this number of predictors, would be demonstrated by the performance of the reconstruction degrading on out-of-sample data.

The results are shown in Fig. 5a, where the mean of 20 000 replications (for each subset size) are shown. This figure shows not only that the in-sample performance (cyan line) improves monotonically, as expected, but also that the predictive performance for withheld data (red line) improves steadily up to the largest testable ensemble size of 8. An



**Fig. 5.** Performance of multiple linear regression as a function of ensemble size. **(a)** Cross-validation using pseudoproxy data from withheld model, with one data point withheld. Lines indicate RMS errors ( $^{\circ}\text{C}$ ) relative to target climate field at location of data points used (cyan line), location of withheld data data points (red line), and global average (blue line). **(b)** Results obtained with real proxy data, showing RMS residuals ( $^{\circ}\text{C}$ ) relative to fitted proxy data (orange), withheld proxy data (green) and estimated RMS errors for true climate field (cyan and red). Dashed cyan and red lines indicate estimated RMS errors relative to true climate field if assumed observational uncertainties are decreased (upper lines) or increased (lower lines) by 20%. All results are means of 20 000 repetitions where 1 data point was withheld (some sampling noise remains).

important additional point to note is that the out of sample performance is substantially better at the data locations, than it is for surface air temperature over the whole globe (blue line). This is partly due to the smoothness of the temperature field and geographical proximity of many data points to each other, but another major reason for this is that the unobserved regions include many of the largest anomalies (such as over the Northern Hemisphere ice sheets), and these tend to be the most highly uncertain between models.

Figure 5b shows that qualitatively similar results are obtained when the real data are used. Here we can only directly assess the performance relative to the noisy data (orange and green lines, for fitted and withheld data respectively). How-

ever, we can estimate the performance relative to the true underlying climate field by subtracting (in quadrature) the observational uncertainties from the actual residuals. These estimated results (red and cyan lines) are however highly sensitive to the assumption that the magnitudes of the observational uncertainties are accurately known, which we discuss further in the following section. It seems optimistic to expect that the errors are determined to within 20% of the correct values (especially since the MARGO uncertainties are only presented in relative terms), and thus quantitative comparison of the performance with real versus pseudoproxy data is challenging. In qualitative terms, however, the results again improve monotonically with ensemble size for fitted and withheld data. Thus, we find no evidence of over-fitting, and retain all 9 models in the multiple regression.

#### 4.2 Data uncertainties and statistical modelling

We can, in principle, attempt to explicitly account for observational uncertainty, by using a weighted regression which accounts for both observational uncertainty, and the system error. That is, rather than minimising the unweighted sum of squared residuals  $\sum_i (s_i - o_i)^2$ , we minimise the weighted sum  $\sum_i [(s_i - o_i)^2 / (\sigma_i^2 + \tau^2)]$  where  $\sigma_i$  are the estimated errors on the respective data points, and  $\tau$  is the system error, that is, the error arising from the inability of the linear combination of models to fit the actual climate field (which must also be estimated).

This approach requires that the errors on the proxy data are well characterised, and furthermore, makes explicit the assumption that the system error is constant in space, at least across the locations of the observations. However, these conditions do not appear to hold. A direct comparison of the 36 points with the largest errors ( $\sigma_i > 3^{\circ}\text{C}$ , with an average value of  $3.7^{\circ}\text{C}$ ) finds that even without any scaling or fitting, all of the models agree rather more closely than this to the observed values, with RMS differences ranging from 2.3 to  $3.0^{\circ}\text{C}$ . Since the models were not tuned to these data, we would expect model errors (i.e. model minus truth) and observational errors (data minus truth) to be independent, adding in quadrature to give model-data differences strictly larger on average than the observational error. Therefore, we conclude that the observational errors of these points are collectively rather smaller than the stated values.

Conversely, if we consider the data points with smaller errors (26 points with  $\sigma_i < 0.8^{\circ}\text{C}$ , averaging  $0.6^{\circ}\text{C}$ ), we find that the untuned models have massively larger RMS residuals at these locations, ranging from 3.8 to  $6.5^{\circ}\text{C}$  across the ensemble, and the results of the multiple linear regression (using all data) only achieves an RMS residual for these points of  $3.8^{\circ}\text{C}$ . More generally, the magnitude of the residuals both for the model fields, and the multiple linear regression results, are actually negatively correlated (albeit to a small degree) with the size of the stated errors. These

difficulties cannot be addressed by varying the scaling of the MARGO reliability index. There is no intrinsic reason why the system error should be negatively correlated with the observational uncertainties, so we propose that another plausible interpretation of these results is that the spread in estimated errors may be due at least in part to methodological differences between the disparate groups of researchers who performed the original analyses over the past few decades. Correlation between errors on different data points (especially among proxy types) is another factor that may play a role, but at present we have no practical means of accounting for this. Therefore, we prefer to use an unweighted fit, which minimises the likelihood that a subset of these data could erroneously bias the result.

However, our results are not sensitive to the use of weighted versus unweighted regression, which may be partly due to the fact that the spread of estimated errors is in fact not very large, with more than 70 % of values lying in the range of 1–2.5 °C.

## 5 Summary and discussion

We have presented a new global reconstruction of SAT and SST for the Last Glacial Maximum. We have shown through extensive cross-validation that the multiple linear regression approach outperforms single model pattern scaling and directly smoothing the proxy data.

Our new estimate of the LGM temperature anomaly is rather warmer than several estimates based on older, less comprehensive, data sets, with our 95 % confidence interval of 3.1–4.7 °C having very little overlap with either of two previous ranges of 4.4–7.2 °C (Schneider von Deimling et al., 2006a, 95 % probability) and 4.6–8.3 °C (Holden et al., 2009, 90 % probability). While part of this discrepancy may be due to methodological differences (in particular the limited ability of intermediate complexity models to adequately represent the spatial pattern of temperature changes), some of it is also due to the fact that the newer proxy data syntheses indicate warmer anomalies than was previously the case, especially over the ocean. For example, the proxy data used here have an unweighted average of 1.6 °C over the tropical ocean (30° S–30° N) and the area mean of our result coincides with this, being  $1.6 \pm 0.7$  °C over the same domain. Earlier estimates favoured rather larger anomalies, for example  $2.7 \pm 1$  °C over the global tropical ocean (Ballantyne et al., 2005) or  $3.0 \pm 0.9$  °C over the tropical Atlantic (Schneider von Deimling et al., 2006b). We note, however, that our reconstruction is still substantially cooler in the tropics than the value of 0.8 °C originally presented by CLIMAP (Climap Project Members, 1976). Some disagreement remains concerning the interpretation of proxy data, especially in the tropical ocean (e.g. de Garidel-Thoron et al., 2007). We have therefore considered the sensitivity of our result to these data, through a simple test in which the observational anoma-

lies for all tropical ocean data points are increased by 1 °C prior to performing the multiple linear regression. With this modification, the mean anomaly in the tropical Atlantic data exceeds the 3.0 °C value used by Schneider von Deimling et al. (2006b). The global temperature anomaly of the resulting reconstruction, however, only increases by 0.3 to 4.3 °C, and even in the tropical region alone, the effect is hardly any greater at 0.4 °C. The resulting reconstruction has a slightly worse fit to the data, with a correlation of only 0.68 versus the original 0.73. Better results might be obtained with a more detailed test that accounts more precisely for the location of different proxy types.

It is perhaps more surprising that our results are rather different from the recent estimate for global SAT anomaly at the LGM of 1.7–3.7 °C (Schmittner et al., 2011, 90 % probability), despite being based on essentially the same data. Many of the PMIP2 models which we used have substantially higher land–sea temperature contrasts than that of the model used in that work, and our reconstruction achieves a notably superior fit to the data, with a correlation of 0.73 between the data and our reconstruction, compared to 0.53 for Schmittner et al. (2011). We therefore consider that our estimate provides a more plausible global interpretation of the proxy data, particularly in reconciling the land and ocean data sets. There is still, however, some indication from the experiments described in Sect. 3.3 that the land–ocean contrast observed in the data is slightly higher than that found in the models. While this could be due to biases in the calibration of the different proxies, other likely causes include the experimental design (with the omission of dust forcing and vegetation feedbacks being obvious candidates due to their likely greater effects over land), or other model inadequacies. In this context, it will be particularly interesting to see whether the forthcoming generation of PMIP3 models can produce a closer fit to the data. However, moderate changes to the pattern of total forcing (and associated response) seem unlikely to lead to major changes in the estimated mean temperature change. A further issue raised by our analysis is the quality of the uncertainty estimates associated with the proxy data. The data which are indicated as having low reliability actually agree rather too well with the model simulations, whereas the reconstruction cannot closely match the data which are considered precise.

One of the major reasons for the intensive study of the LGM is the hope that it might help us to better constrain the equilibrium climate sensitivity (and perhaps also other climatic changes), due to the large and reasonably well-constrained forcing and temperature response. A first order estimate of the equilibrium climate sensitivity could be generated from the ratio of temperature change to radiative forcing. Our new temperature anomaly of  $4.0 \pm 0.8$  °C, combined with estimated forcing of 6–11 W m<sup>-2</sup> (Annan et al., 2005; Jansen et al., 2007) would suggest a median estimate for the equilibrium climate sensitivity of around 1.7 °C, with a 95 % range of 1.2–2.4 °C. However, such a simplistic estimate is

far from robust, as it ignores any asymmetry or nonlinearity which is thought to exist in the response to different forcings (Hargreaves et al., 2007; Yoshimori et al., 2011). The ratio between temperature anomalies obtained under LGM and doubled CO<sub>2</sub> conditions found in previous modelling studies varies from 1.3 (Schmittner et al., 2011) to over 2 (Schneider von Deimling et al., 2006a). More recently, Hargreaves et al. (2012) used the relationship found in the PMIP2 ensemble between the tropical temperature change at the LGM, and equilibrium climate sensitivity, to estimate the equilibrium climate sensitivity to be around 2.5 °C with a high probability of lying under 4 °C, although this result is subject to several important caveats.

Understanding and quantifying the relationship between past and future climate changes remains a major challenge, but our robust estimate of temperature change at the LGM, based on current understanding of proxy data, is an important step towards this goal.

**Supplementary material related to this article is available online at: <http://www.clim-past.net/9/367/2013/cp-9-367-2013-supplement.zip>.**

*Acknowledgements.* We are grateful for helpful comments provided by Andreas Schmittner and Thomas Schneider von Deimling. We acknowledge the international modelling groups for providing their data for analysis, and the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) for collecting and archiving the model data. The PMIP 2 Data Archive is supported by CEA, CNRS and the Programme National d'Etude de la Dynamique du Climat (PNEDC). We are also very grateful to the authors of MARGO Project Members (2009), Bartlein et al. (2011) and Shakun et al. (2012) for making their analyses available. This research was supported by the Environment Research and Technology Development Fund (S-10-3) of the Ministry of the Environment, Japan.

Edited by: V. Rath

## References

- Abe-Ouchi, A., Segawa, T., and Saito, F.: Climatic Conditions for modelling the Northern Hemisphere ice sheets throughout the ice age cycle, *Clim. Past*, 3, 423–438, doi:10.5194/cp-3-423-2007, 2007.
- Annan, J. D., Hargreaves, J. C., Ohgaito, R., Abe-Ouchi, A., and Emori, S.: Efficiently constraining climate sensitivity with paleoclimate simulations, *Scientific Online Letters on the Atmosphere*, 1, 181–184, 2005.
- Ballantyne, A. P., Lavine, M., Crowley, T. J., Liu, J., and Baker, P. B.: Meta-analysis of tropical surface temperatures during the Last Glacial Maximum, *Geophys. Res. Lett.*, 32, 5, doi:10.1029/2004GL021217, 2005.
- Bartlein, P., Harrison, S., Brewer, S., Connor, S., Davis, B., Gajewski, K., Guiot, J., Harrison-Prentice, T., Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppä, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A. E., Williams, J., and Wu, H.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, *Clim. Dynam.*, 37, 775–802, 2011.
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, Th., Hewitt, C. D., Kageyama, M., Kitoh, A., Laine, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., Yu, Y., and Zhao, Y.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features, *Clim. Past*, 3, 261–277, doi:10.5194/cp-3-261-2007, 2007.
- Climap Project Members: The surface of the ice-age earth, *Science*, 191, 1131–1137, 1976.
- Crowley, T.: CLIMAP SSTs re-visited, *Clim. Dynam.*, 16, 241–255, 2000.
- Crucifix, M. and Hewitt, C. D.: Impact of vegetation changes on the dynamics of the atmosphere at the Last Glacial Maximum, *Clim. Dynam.*, 25, 447–459, 2005.
- de Garidel-Thoron, T., Rosenthal, Y., Beaufort, L., Bard, E., Sonzogni, C., and Mix, A.: A multiproxy assessment of the western equatorial Pacific hydrography during the last 30 kyr, *Paleoceanography*, 22, PA3204, 2007.
- Hansen, J. and Lebedeff, S.: Global trends of measured surface air temperature, *J. Geophys. Res.*, 92, 13345–13372, 1987.
- Hargreaves, J. C., Abe-Ouchi, A., and Annan, J. D.: Linking glacial and future climates through an ensemble of GCM simulations, *Clim. Past*, 3, 77–87, doi:10.5194/cp-3-77-2007, 2007.
- Hargreaves, J. C., Paul, A., Ohgaito, R., Abe-Ouchi, A., and Annan, J. D.: Are paleoclimate model ensembles consistent with the MARGO data synthesis?, *Clim. Past*, 7, 917–933, doi:10.5194/cp-7-917-2011, 2011.
- Hargreaves, J. C., Annan, J. D., Yoshimori, M., and Abe-Ouchi, A.: Can the Last Glacial Maximum constrain climate sensitivity?, *Geophys. Res. Lett.*, 39, 24, doi:10.1029/2012GL053872, 2012.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C.: Bayesian model averaging: A tutorial, *Stat. Sci.*, 14, 382–401, 1999.
- Holden, P. B., Edwards, N. R., Oliver, K. I. C., Lenton, T. M., and Wilkinson, R. D.: A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1, *Clim. Dynam.*, 35, 1–22, doi:10.1007/s00382-009-0630-8, 2009.
- Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D.: Palaeoclimate, 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*, chap. 6, Cambridge University Press, Cambridge, UK and New York, NY, USA., 2007.
- Krishnamurti, T., Kishtawal, C., Zhang, Z., LaRow, T., Bachiochi, D., Williford, E., Gadgil, S., and Surendran, S.: Multimodel ensemble forecasts for weather and seasonal climate, *J. Climate*, 13, 4196–4216, 2000.
- MARGO Project Members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nat. Geosci.*, 2, 127–132, doi:10.1038/NGEO411, 2009.
- Santer, B., Wigley, T., Schlesinger, M., and Mitchell, J.: Developing climate scenarios from equilibrium GCM results, *Tech. Rep. Technical Note 47*, Max Planck Institut für

- Meteorologie, Hamburg, 1990.
- Schmittner, A., Urban, N., Shakun, J., Mahowald, N., Clark, P., Bartlein, P., Mix, A., and Rosell-Melé, A.: Climate Sensitivity Estimated from Temperature Reconstructions of the Last Glacial Maximum, *Science*, 334, 1385–1388, 2011.
- Schneider von Deimling, T., Ganopolski, A., Held, H., and Rahmstorf, S.: How cold was the Last Glacial Maximum?, *Geophys. Res. Lett.*, 33, L14709, doi:10.1029/2006GL026484, 2006a.
- Schneider von Deimling, T., Held, H., Ganopolski, A., and Rahmstorf, S.: Climate sensitivity estimated from ensemble simulations of glacial climate, *Clim. Dynam.*, 27, 149–163, 2006b.
- Shakun, J., Clark, P., He, F., Marcott, S., Mix, A., Liu, Z., Otto-Bliesner, B., Schmittner, A., and Bard, E.: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation, *Nature*, 484, 49–54, 2012.
- Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), *J. Climate*, 21, 2283–2296, doi:10.1175/2007JCLI2100.1, 2008.
- Snedecor, G. and Cochran, W.: *Statistical Methods*, 8th Edn., Iowa state University press, Ames, Iowa, 1989.
- Yoshimori, M., Hargreaves, J., Annan, J., Yokohata, T., and Abe-Ouchi, A.: Dependency of Feedbacks on Forcing and Climate State in Perturbed Parameter Ensembles, *J. Climate*, 24, 6440–6455, 2011.