

This discussion paper is/has been under review for the journal Climate of the Past (CP).
Please refer to the corresponding final paper in CP if available.

Technical Note: Are large error bars desirable? A note on quantitative model-proxy comparison

J. Liakka^{1,2}, J. T. Eronen^{1,2}, H. Tang³, and F. T. Portmann^{1,2,4}

¹Biodiversity and Climate Research Centre (LOEWE BiK-F), Frankfurt, Germany

²Senckenberg Gesellschaft für Naturforschung, Frankfurt, Germany

³Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

⁴Institute of Physical Geography, Goethe University, Frankfurt, Germany

Received: 11 November 2014 – Accepted: 15 November 2014

– Published: 15 December 2014

Correspondence to: J. Liakka (johan.liakka@senckenberg.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

The combined use of proxy records and climate modelling is invaluable for obtaining a better understanding of past climates. However, many methods of model-proxy comparison in the literature are fundamentally problematic because larger errors in the proxy tend to yield a “better” agreement with the model. Here we quantify model-proxy agreement as a function to proxy uncertainty using the overlapping coefficient OVL, which measures the similarity between two probability distributions. We found that the model-proxy agreement is poor (OVL < 50%) if the proxy uncertainty (σ_p) is greater than three times the model variability (σ_m), *even if* the model and proxy have similar mean estimates. Hence only proxies that fulfil the condition $\sigma_p < 3\sigma_m$ should be used for detailed quantitative evaluation of the model performance.

1 Introduction

In paleoclimatology, the combined use of proxy records (reconstruction of past climate conditions) and climate modelling is invaluable for understanding past climate dynamics. Comparing climate model simulations with proxy records is also useful for evaluating the model performance. If the models perform well in the past, this can eventually increase the reliability of future climate projections (Braconnot et al., 2012; Schmidt et al., 2014). Compared to modern observations, proxy records are more uncertain and many times qualitative. Therefore, the model-proxy comparison is often performed purely visually or qualitatively (e.g., Jost et al., 2005; Lunt et al., 2008; Braconnot et al., 2012; Jiang et al., 2012, 2013). Quantitative model-proxy comparisons have been used more in recent years as quantitative proxy reconstructions have become increasingly available. However, these quantitative reconstructions are often associated with large uncertainties and various distributions, thus simply comparing the mean values and SDs between the model and proxy similar to that for the present day (Taylor, 2001; Gleckler et al., 2008; Hargreaves et al., 2013) is questionable.

CPD

10, 4535–4552, 2014

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Are large error bars desirable?

J. Liakka et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

To take the uncertainties of proxy data into account, several methods have been proposed to evaluate the (dis)agreement between the model and proxy data. The simplest one is to use the median and interquartile/total range of the proxy and model data to check whether they overlap with each other (e.g., Steppuhn et al., 2007; Otto-Bliesner et al., 2009; Micheels et al., 2011; Bradshaw et al., 2012; Lunt et al., 2012; Harrison et al., 2014). One major problem with such visual evaluation and simple test of overlap is that a larger uncertainty in the proxy record is preferable to get a good agreement with the model. Increasing the proxy uncertainty essentially increases the probability that the modelled value lies within the error bars of the proxy. A more sophisticated method was introduced by Guiot et al. (1999), who used the so-called “fuzzy distance” to measure the difference between model and proxy data. In their method, increasing uncertainties in the model or proxy data can either increase or decrease the fuzzy distance depending on how the data is distributed. Although sophisticated, this method has not been extensively used by the community.

In this study we elucidate some of the (potentially serious) issues associated with the conventional model-proxy comparison methods. Since many of these methods yield an improved model-proxy agreement for large uncertainties in the proxy data, we dedicate particular attention on this feature. To quantify agreement we calculate the overlapping coefficient (OVL), which measures the degree of similarity between two probability distributions (Inman and Bradley Jr., 1989). To create these probability distributions for the model and proxy data, the model variability as well as proxy uncertainty/variability (hereafter referred to as “proxy uncertainty” for simplicity) need to be accounted for. In the next section, the theoretical background of OVL is discussed in more detail. In Sect. 3, we apply OVL to some simple conceptual cases as well as to a real example consisting of a climate model experiment of the Late Miocene. This is followed by our conclusions in Sect. 4.

2 Method for evaluating model-proxy agreement

2.1 Theoretical background

The method for model-proxy comparison introduced here is based on a probabilistic approach. If $f_1(x)$ and $f_2(x)$ are two probability (density) functions of a certain variable x , one can measure the degree of similarity between these two distributions by calculating the overlapping coefficient OVL (e.g., Inman and Bradley Jr., 1989)

$$\text{OVL} = \int_{-\infty}^{\infty} \min(f_1(x), f_2(x)) dx. \quad (1)$$

Hence, OVL yields the degree of overlapping between f_1 and f_2 , or in other words, the degree of similarity between the two distributions. Here f_1 and f_2 represent the probability distributions from climate models (hereafter referred to with subscript m: f_m) and proxies (subscript p: f_p), and x some climate variable. Since OVL is a probabilistic measure it is expressed in %, and thus varies between 0% (no agreement) and 100% (complete agreement). For simplicity, we use OVL = 50% as a threshold for satisfactory agreement. Note that OVL refers to agreement and not to any statistical probability. These are, however, somewhat related: if the SD of the proxy data is greater than that of the model, OVL is similar to the probability that the true value in the proxy is captured by the model \pm two SDs (Fig. S1 in the Supplement).

A major advantage of using OVL over other probabilistic measures is its simplicity; Eq. (1) can be applied to any model and proxy data once f_m and f_p are known. Prior to calculating OVL, however, one must ensure that f_m and f_p are probability distributions of a variable, which is represented at the same point in time and space. Hence, OVL should essentially be used for point-by-point comparisons only. The drawback of the OVL method is that it does not provide any information on the (dis)agreement expressed in data units (e.g. in $^{\circ}\text{C}$ or mm yr^{-1}). Therefore OVL should preferably be

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



used together with some measure of the distance between the model and proxy, e.g. the difference between the mean estimates.

2.2 Distributions of modelled and observed data

The main advantage of the overlapping (OVL) method is that the proxy uncertainty and model variability are accounted for as they alter the SDs of the corresponding probability distributions. A necessary requirement for calculating OVL (Eq. 1) is that one can estimate the probability distributions of both the model and the proxy data. For this, it is necessary to have sufficient data to calculate data-specific statistical quantities such as the mean (\bar{x}) and the SD (σ). Second, it is important to fit appropriate probability distributions to the data.

In the following we discuss probability distributions derived from climate models and proxy data. We restrict our discussion to examples from our previous work, which includes vegetation proxies based on the co-existence approach of the mean annual temperature (MAT) and precipitation (MAP) (Utescher et al., 2014), as well as mammal proxies of MAP (Eronen et al., 2010; Liu et al., 2012). Note, however, that the methodology provided here is also applicable to other quantitative proxy sources.

In climate models, temperature and precipitation variability usually follow a Gaussian (Normal) and a gamma distribution, respectively. For both distributions, only \bar{x}_m and σ_m are required to calculate the corresponding probability functions. For simplicity, we define σ_m as the “inter-variability” of x ; for example, when \bar{x}_m is the annual mean, σ_m represents the inter-annual variability. Since the climate variability recorded in the proxy is often associated with lower frequency, it is likely that inter-annual variations in the model underestimate the variability of the proxy data. In proxy records it is however often difficult to separate between uncertainty from variability. Consequently, assuming inter-annual variability is sufficient to understand the general behaviour of OVL.

The use of a gamma instead of a Gaussian distribution for precipitation is motivated by the fact that the probability frequency is zero for negative values, which is one of the main features of the gamma distribution. Furthermore, the gamma distribution can

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Are large error bars desirable?

J. Liakka et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

potentially change the shape depending on the characteristics of the data. If $\bar{x} < \sigma$, the gamma distribution is exponential with decreasing probabilities toward higher values (e.g., Fig. 1 in Husak et al., 2007). If, on the other hand $\bar{x} > \sigma$, the gamma distribution is zero at $x = 0$ and has a skewed “Gaussian-like” shape (Fig. 1 in Husak et al., 2007).

5 The advantage of using the gamma distribution to represent precipitation frequency has been noted in several studies, e.g., Geng et al. (1986), Watterson and Dix (2003) and Husak et al. (2007).

Vegetation proxies based on the co-existence approach yield intervals with homogeneous probability rather than a most likely estimate and a standard error (Utescher et al., 2014). Hence, this proxy is represented by a distribution with equal probability for all values within the co-existence range $[x_{\min}, x_{\max}]$:

$$f(x) = \begin{cases} 1/(x_{\max} - x_{\min}), & x_{\min} \leq x \leq x_{\max} \\ 0, & \text{elsewhere.} \end{cases} \quad (2)$$

Using mammal proxies to estimate past MAP yields an estimate of the mean along with a standard error of 388 mm (Eronen et al., 2010). To avoid negative precipitation values, mammal data is best represented by a gamma distribution.

3 Results

3.1 Some simple cases

To better understand the behavior of OVL, Fig. 1 shows some illustrative examples. Figure 1a shows an example of the homogeneous probability distribution that we use for the vegetation proxy (see Eq. 2). Thus, when comparing vegetation proxies with model data, one must deal with two different probability distributions. As a consequence, the maximum OVL value cannot be as high as 100%, but will depend on the properties of the corresponding distributions. Figure 1a depicts the maximum

possible agreement between the homogeneous distribution and a Gaussian distribution ($\approx 82\%$); compared with a gamma distribution instead, the maximum agreement is very similar to the Gaussian case ($\text{OVL} \approx 81\%$; not shown).

In the next two panels (Fig. 1b and c) we consider OVL of two gamma distributions.

The “model” distribution is assumed to be the same in both cases, but the “proxy” distribution differs significantly; in the first example (b), the proxy has the same variability as the model, but its mean value is shifted so that the mean value of the model is clearly situated outside the error bars of the proxy. In the second example (c), the proxy and model have the same mean, but the proxy error is $4\times$ the model variability (i.e. $\sigma_p = 4\sigma_m$). Despite having the same mean value in (c), our results suggest that the model-proxy agreement is worse than in (b) ($\text{OVL} = 40\%$ vs. $\text{OVL} = 45\%$).

The poor agreement in Fig. 1c is associated with the large uncertainty of the proxy. To reveal a better understanding on how proxy uncertainty ranges influence OVL, we derive a simple analytical relationship between OVL and (proxy) uncertainty based on two hypothetical probability distributions with triangular geometry (see Appendix A and Fig. S2). If both of these triangular distributions have the same mean ($\Delta\bar{x} = 0$ in Fig. S2), the overlapping coefficient is simply given by (see Appendix A for derivation):

$$\text{OVL} = \frac{2}{1+q}, \quad (3)$$

where q is a non-dimensional parameter of the length ratio (i.e. SD ratio) of the wider triangle to the narrower. Thus, assuming that the proxy uncertainty is greater than the model variability, q is simply given by $q \equiv \sigma_p/\sigma_m$. If we apply the simple analytical solution in Eq. (3) to the example in Fig. 1c we obtain $\text{OVL} = 40\%$ ($q = 4 \Rightarrow \text{OVL} = 2/(1+4) = 40\%$), thus exactly the same as in the case with two gamma distributions in Fig. 1c. In fact, despite its simplicity, the analytical solution shows a very good agreement with other conventional probability distributions for the whole q spectrum (Fig. 2a). As a result, if the mean values from the model and proxy are similar, $\text{OVL} > 50\%$ requires that $q < 3$, i.e. that the proxy uncertainty is less than $3\times$

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the model variability. In other words, if $q > 3$ the model and proxy are most probably *not* in agreement even if their means are similar. Thus, proxies with such large uncertainty should preferably not be used for a quantitative evaluation of the model performance.

Somewhat surprisingly, however, is that large uncertainty proxies in some cases have a greater agreement with the model if their means are further apart. This is illustrated in Fig. 2b, which shows OVL as a function of q with the proxy mean located at $\bar{x}_p = \bar{x}_m + 2\sigma_m$. In particular for the exponential shape of the gamma distribution, there is a range of q values between ~ 2 and ~ 6.5 for which OVL is greater than 50%. Hence, for proxies that are best represented by an exponential gamma distribution, e.g. low precipitation estimates with a large uncertainty, the agreement with the model becomes *better* if the model mean is *significantly lower* than the proxy mean. In this special case, similar mean estimates are only desirable for a good agreement if the proxy uncertainty is within the same order of magnitude as the model variability.

3.2 Application on real data

In this section, we apply the OVL approach to real climate model and proxy data. For this purpose we use data from a previously published paleoclimate modelling study (Micheels et al., 2011). Micheels et al. (2011) conducted a fully-coupled atmosphere–ocean simulation of the late Miocene (7–11 millions of years ago) using ECHAM5/MPI-OM, where ECHAM5 is the atmosphere model (e.g. Roeckner et al., 2006) and MPI-OM the ocean model (e.g. Jungclaus et al., 2006). In Micheels et al. (2011), ECHAM5/MPI-OM was integrated to equilibrium (2500 years) with late Miocene boundary conditions. The late Miocene climatology was constructed from the last 10 years of the 2500-year simulation. These results were then compared to late Miocene estimates of MAT and MAP from vegetation fossils (co-existence approach) as well as MAP from mammal hypsodonty. In total, the proxy dataset consists of 472 fossil site records with 69 (60) MAT (MAP) estimates derived from vegetation and 343 MAP estimates from mammals, respectively. The proxy dataset is the same as in Micheels et al. (2011) except for a few additional vegetation fossils from Siberia (Popova et al.,

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2012). Therefore the reader is referred to Micheels et al. (2011) for more details on the proxy records and model results. In this study we discuss only the methodology of the model-proxy comparison.

To evaluate the model-proxy agreement Micheels et al. (2011) used a method that was first introduced by Steppuhn et al. (2007). This method considers the minimum distance between the climate ranges of the model and proxy. If this distance is zero, i.e. the error bars of the proxy overlap with the model variability, the model and proxy are considered to be in agreement. If there is no overlap, the model-proxy (dis)agreement is quantified by the minimum distance between the climate intervals. Hereafter we refer to this method as the “minimum distance method”. For the vegetation proxy these intervals are constructed as a result of the co-existence approach, which yields a maximum and a minimum estimate. For the mammal proxy, the intervals are constructed by taking the mean MAP estimate \pm the standard error of 388 mm. If the mean estimate is less than 388 mm, the lower bound of the interval is set to zero. The climate intervals in the model were constructed by taking the maximum and minimum values of MAT and MAP for each gridpoint from the last 10 yr of the model simulation.

Using the minimum distance method, the modelled MAT agrees fairly well with the MAT from the vegetation proxy (Fig. 3a). The best agreement is achieved over the central parts of Europe and Asia. North (South) of these regions the model is generally too cold (warm). In terms of MAP, the model-proxy agreement seems to be even better (Fig. 3b), especially when comparing with the mammal data. With respect to the vegetation proxy, the model has a good agreement over most of Asia, but produces too dry conditions over most of Europe. In essence, the good agreement between the model and the mammal proxy in Fig. 3b results from the fact that the modelled MAP over most localities is within the interval given by the MAP estimate from mammals \pm the standard error (388 mm).

Figure 3c and d show the corresponding values of OVL for the MAT and MAP cases, respectively. The grey markers depict the cases for which $OVL > 50\%$. According to the OVL calculations the MAT agreement over central Europe remains fairly high, whereas

Are large error bars desirable?

J. Liakka et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

the agreement over central Asia vanishes almost completely as compared with the minimum distance method. In terms of MAP, the difference is even more pronounced; for most of the localities that exhibited a model-proxy agreement with the minimum distance method, OVL is substantially lower than 50%.

To reveal some insight into the low OVL, Fig. 3e and f display the corresponding values of q ($= \sigma_p / \sigma_m$) for all localities. Above we found that similar model and proxy means can yield $OVL > 50\%$ only if $q < 3$. For most vegetation proxies, q is less than 3 (Fig. 3e and circles in f). Hence, for these records the relatively poor OVL agreement is caused by different mean values between the model and the proxy. The large uncertainty of the mammal data, however, yields mostly high q values, especially over southern Europe as well as western and central Asia (triangles in Fig. 3f). Thus, the poor OVL agreement in these regions can be explained by large uncertainties in the proxy. Further, the somewhat better agreement over southeast Asia coincides with much lower q values in that region. Note that there are a couple of localities that exhibit $OVL > 50\%$ despite high q values (e.g. on the Arabic peninsula). These localities are predominately associated with low precipitation rates, i.e. an exponential shape of the gamma distribution. Hence, for these localities the situation discussed in Fig. 2b applies: despite a high q value, OVL is large because the model mean is significantly smaller than the mean estimate from the proxy (dashed line in Fig. 2b).

4 Conclusions

This work was motivated by the fact that many conventional model-proxy comparisons favour a good agreement for large errors in the proxy. These methods are fundamentally problematic because the model “performance” is largely determined by the data used for comparison. Here we illustrate how uncertainty of the proxy influences the model-proxy agreement. We use a simple metric called the overlapping coefficient (OVL), which measures agreement of two probability distributions. Even if OVL has some shortcomings, it has the ability to quantify agreement as a function of uncertainty.

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Our main result is that the model-proxy agreement can be poor even if the mean values are similar. More specifically, for similar means OVL is always less than 50 % if the proxy uncertainty is greater than 3× the model variability (i.e. $q > 3$). We can use this result to distinguish between disagreement due to bad model performance (difference in means) and uncertainty of the proxy. For localities that exhibit low OVL values and $q < 3$, the poor agreement is attributed to the model mean being too far apart from the proxy mean. In case of $q > 3$ the model and proxy could be in agreement, but the uncertainty of the proxy is too large to draw any such conclusions. Most importantly, this result shows that proxies with such large uncertainties should be avoided for detailed quantitative model-proxy comparisons.

Appendix A: Analytical relationship between OVL and uncertainty

Here we examine how uncertainty in (proxy) data, i.e. the magnitude of error bars influences OVL for a given (model) distribution. Our analysis here is based on two hypothetical triangular probability distributions (Fig. S2). The advantage of using such a simple geometry is that we can derive analytical relationships. For simplicity, we define the width of each triangle to be equal to two SDs (2σ). To ensure that the area of each triangle is one, the maximum height is simply given by $1/\sigma$. For simplicity, we assume that both distributions have the same mean, i.e. $\Delta\bar{x} = 0$ in Fig. S2, and that $\sigma_2 \geq \sigma_1$. With these simplifications the area common to both distributions, i.e. the overlapping coefficient OVL, is given by simple geometry:

$$\text{OVL} = 1 - f_j(\sigma_2 - \sigma_1), \quad (\text{A1})$$

where f_j is the probability density at which the triangles intersect. Because both triangles have the same mean, the intersections occur at the same f value on both sides of the distribution center. Simple algebra yields $f_j = 1/(\sigma_1 + \sigma_2)$, implying that

$$\text{OVL} = \frac{2\sigma_1}{\sigma_1 + \sigma_2}. \quad (\text{A2})$$

Defining q as the ratio of the SDs ($q \equiv \sigma_2/\sigma_1$) entails

$$\text{OVL} = \frac{2}{1+q}, \quad (\text{A3})$$

which is equal to Eq. (3). The solution of Eq. (A3) is shown in Fig. 2a (grey solid line). OVL decreases from 100 % at $q = 1$ to about 10 % at $q = 20$. The decline is much faster for smaller values of q : already for $q > 3$ is $\text{OVL} < 50\%$. Although using simplified geometry to arrive at Eq. (A3), its solution is very similar to some other commonly used probability distributions (Fig. 2a).

**The Supplement related to this article is available online at
doi:10.5194/cpd-10-4535-2014-supplement.**

Acknowledgements. This work was supported by the LOEWE research funding programme of the state of Hesse, the German Research Foundation (grant MI 926/8-1), and from the Marie Curie Programme of the European Commission. We are indebted to Arne Micheels and Torsten Utescher for providing the climate modelling and paleovegetation data. Please contact the corresponding or second author to obtain the data that was used to produce Fig. 3. We thank Liam Langan and Robert O'Hara (BiK-F) for fruitful discussions on the manuscript.

References

- Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, *Nature Climate Change*, 2, 417–424, 2012. 4536
- Bradshaw, C. D., Lunt, D. J., Flecker, R., Salzmann, U., Pound, M. J., Haywood, A. M., and Eronen, J. T.: The relative roles of CO_2 and palaeogeography in determining late Miocene climate: results from a terrestrial model–data comparison, *Clim. Past*, 8, 1257–1285, doi:10.5194/cp-8-1257-2012, 2012. 4537
- Eronen, J., Puolamäki, K., Liu, L., Lintulaakso, K., Damuth, J., Janis, C., and Fortelius, M.: Precipitation and large herbivorous mammals, Part I: Estimates from present-day communities, *Evol. Ecol. Res.*, 12, 217–233, 2010. 4539, 4540

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Geng, S., Penning de Vries, F. W., and Supit, I.: A simple method for generating daily rainfall data, *Agr. Forest Meteorol.*, 36, 363–376, 1986. 4540
- Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, *J. Geophys. Res.*, 113, D06104, doi:10.1029/2007JD008972, 2008. 4536
- 5 Guiot, Boreux, J. J., Braconnot, P., and Torre, F.: Data-model comparison using fuzzy logic in paleoclimatology, *Clim. Dynam.*, 15, 569–581, 1999. 4537
- Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A.: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene, *Clim. Past*, 9, 811–823, doi:10.5194/cp-9-811-2013, 2013. 4536
- 10 Harrison, S., Bartlein, P., Brewer, S., Prentice, I., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, *Clim. Dynam.*, 43, 671–688, 2014. 4537
- Husak, G. J., Michaelsen, J., and Funk, C.: Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications, *Int. J. Climatol.*, 27, 935–944, 2007. 4540
- 15 Inman, H. F. and Bradley Jr., E. L.: The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities, *Commun. Stat. Theory*, 18, 3851–3874, 1989. 4537, 4538
- Jiang, D., Lang, X., Tian, Z., and Wang, T.: Considerable model-data mismatch in temperature over China during the mid-Holocene: Results of PMIP simulations, *J. Climate*, 25, 4135–4153, 2012. 4536
- 20 Jiang, D., Tian, Z., and Lang, X.: Mid-Holocene net precipitation changes over China: model-data comparison, *Quaternary Sci. Rev.*, 82, 104–120, 2013. 4536
- Jost, A., Lunt, D., Kageyama, M., Abe-Ouchi, A., Peyron, O., Valdes, P., and Ramstein, G.: High-resolution simulations of the last glacial maximum climate over Europe: a solution to discrepancies with continental palaeoclimatic reconstructions?, *Clim. Dynam.*, 24, 577–590, 2005. 4536
- 25 Jungclaus, J., Keenlyside, N., Botzet, M., Haak, H., Luo, J.-J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM, *J. Climate*, 19, 3952–3972, 2006. 4542
- 30 Liu, L., Puolamäki, K., Eronen, J., Mirzaie, M., Hernesniemi, E., and Fortelius, M.: Estimating net primary productivity from dental traits, *P. R. Soc. B*, 279, 2793–2799, 2012. 4539

Are large error bars desirable?

J. Liakka et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

- Lunt, D. J., Flecker, R., Valdes, P. J., Salzmann, U., Gladstone, R., and Haywood, A. M.: A methodology for targeting palaeo proxy data acquisition: a case study for the terrestrial late Miocene, *Earth Planet. Sc. Lett.*, 271, 53–62, 2008. 4536
- Lunt, D. J., Dunkley Jones, T., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., Loptson, C., Marotzke, J., Roberts, C. D., Tindall, J., Valdes, P., and Winguth, C.: A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP, *Clim. Past*, 8, 1717–1736, doi:10.5194/cp-8-1717-2012, 2012. 4537
- Micheels, A., Bruch, A. A., Eronen, J., Fortelius, M., Harzhauser, M., Utescher, T., and Mosbrugger, V.: Analysis of heat transport mechanisms from a Late Miocene model experiment with a fully-coupled atmosphere–ocean general circulation model, *Palaeogeogr. Palaeoclimatol.*, 304, 337–350, 2011. 4537, 4542, 4543, 4552
- Otto-Bliesner, B. L., Schneider, R., Brady, E., Kucera, M., Abe-Ouchi, A., Bard, E., Braconnot, P., Crucifix, M., Hewitt, C., Kageyama, M., Marti, O., Paul, A., Rosell-Melé, A., Waelbroeck, C., Weber, S. L., Weinelt, M., and Yu, Y.: A comparison of PMIP2 model simulations and the MARGO proxy reconstruction for tropical sea surface temperatures at last glacial maximum, *Clim. Dynam.*, 32, 799–815, 2009. 4537
- Popova, S., Utescher, T., Gromyko, D., Bruch, A. A., and Mosbrugger, V.: Palaeoclimate evolution in Siberia and the Russian Far East from the Oligocene to Pliocene: evidence from fruit and seed floras, *Turk. J. Earth Sci.*, 21, 315–334, 2012. 4542
- Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, *J. Climate*, 19, 3771–3791, 2006. 4542
- Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L.-B., and Yiou, P.: Using palaeo-climate comparisons to constrain future projections in CMIP5, *Clim. Past*, 10, 221–250, doi:10.5194/cp-10-221-2014, 2014. 4536
- Stappuhn, A., Micheels, A., Bruch, A. A., Uhl, D., Utescher, T., and Mosbrugger, V.: The sensitivity of ECHAM4/ML to a double CO₂ scenario for the late Miocene and the comparison to terrestrial proxy data, *Global Planet. Change*, 57, 189–212, 2007. 4537, 4543, 4552
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, 2001. 4536

Utescher, T., Bruch, A., Erdei, B., François, L., Ivanov, D., Jacques, F., Kern, A., Liu, Y.-S., Mosbrugger, V., and Spicer, R.: The coexistence approach – theoretical background and practical considerations of using plant fossils for climate quantification, *Palaeogeogr. Palaeoclimatol.*, 410, 58–73, 2014. 4539, 4540

- 5 Watterson, I. and Dix, M.: Simulated changes due to global warming in daily precipitation means and extremes and their interpretation using the gamma distribution, *J. Geophys. Res.*, 108, 4379, doi:10.1029/2002JD002928, 2003. 4540

Are large error bars desirable?

J. Liakka et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Are large error bars desirable?

J. Liakka et al.

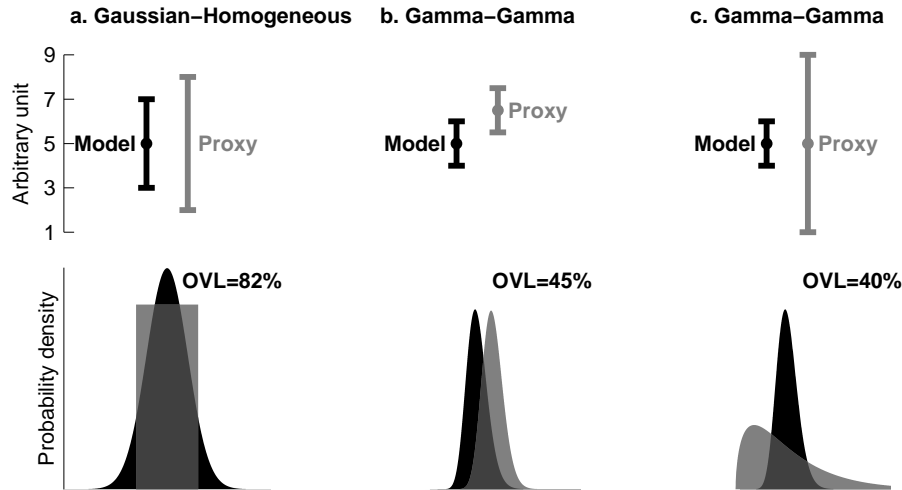


Figure 1. Three examples of different model and proxy distributions, and their corresponding overlapping coefficient OVL computed by Eq. (1).

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	⏩
⏴	⏵
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Are large error bars desirable?

J. Liakka et al.

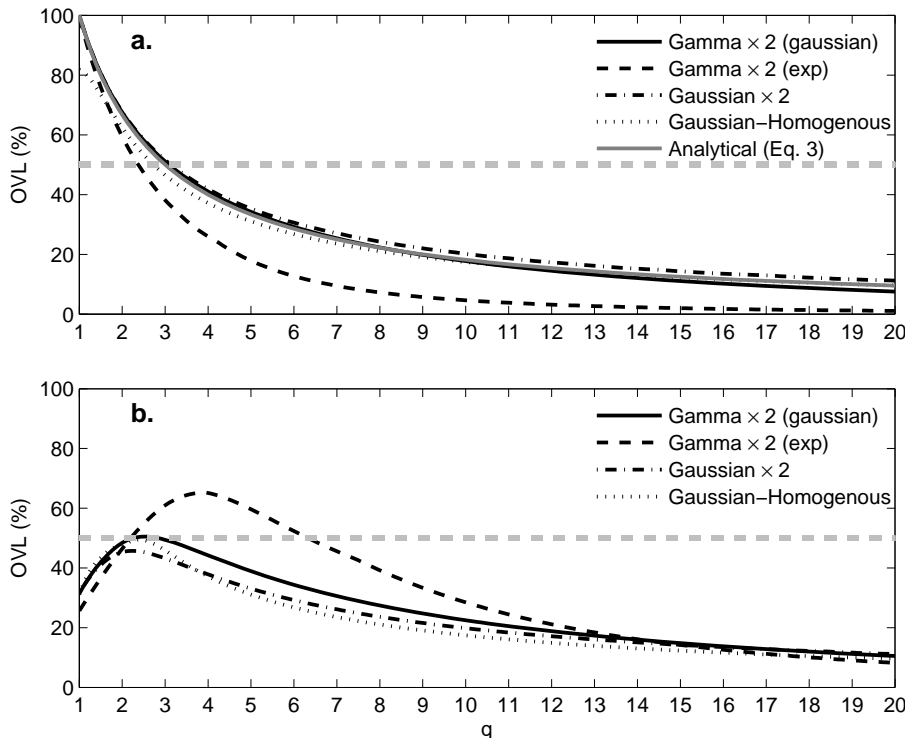


Figure 2. OVL as a function of q ($= \sigma_p / \sigma_m$) for the distributions given by the legend. The “gaussian” and “exp” in the legend refers, respectively, to a “Gaussian-like” and an exponential shape of the gamma distribution. In the upper panel (a) the mean values of the model and proxy are assumed to be the same ($\bar{x}_p = \bar{x}_m$). The grey solid line represents the analytical solution derived from triangular geometry (Eq. 3). In (b) the mean of the proxy is shifted so that $\bar{x}_p = \bar{x}_m + 2\sigma_m$. The SD of the homogeneous distribution is defined as $(x_{\max} - x_{\min})/3$ so that the maximum OVL is obtained for $q = 1$ (cf. Fig. 1a). The horizontal grey dashed line depicts OVL = 50%, which we use a threshold for satisfactory agreement in our study.

[Title Page](#)

Abstract	Introduction
Conclusions	References
Tables	Figures

[◀](#)
[▶](#)

[◀](#)
[▶](#)

Back	Close
----------------------	-----------------------

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Are large error bars desirable?

J. Liakka et al.

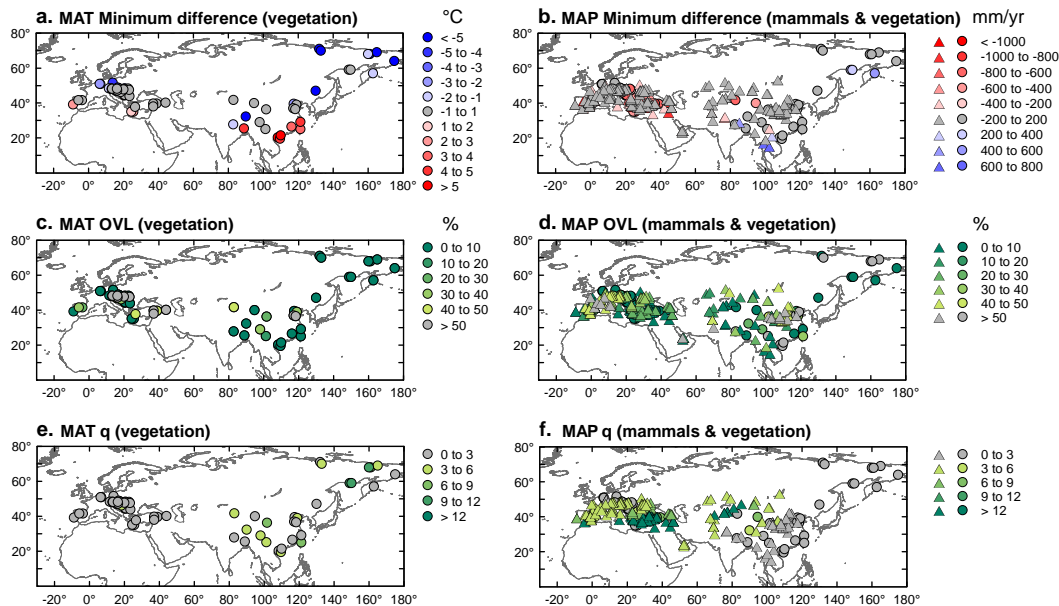


Figure 3. Comparison of MAT (**a**, **c** and **e**) and MAP (**b**, **d** and **f**) between the simulations and proxies from Micheels et al. (2011) using the minimum distance method (Steppuhn et al., 2007) (**a** and **b**), and OVL (**c** and **d**). Panels (**e**) and (**f**) show the corresponding values of q ($= \sigma_p / \sigma_m$). Vegetation proxies are displayed as circles whereas mammal proxies are depicted as triangles.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

