Calculating uncertainties on predictions of palaeoprecipitation from the magnetic properties of soils

David Heslop *, Andrew P. Roberts

Research School of Earth Sciences, Australian National University, 0200 Canberra, ACT, Australia

A R T I C L E   I N F O

Article history:
Received 16 July 2012
Accepted 27 November 2012
Available online xxxx

Keywords:
climofunction
discrimination intervals
palaeoprecipitation
pedogenesis
magnetic susceptibility

A B S T R A C T

Quantitative predictions of past climate states based on calibrated proxy data are key to the reconstruction of palaeoenvironments and are essential for climate model validation. Magnetic climofunctions have been used to make predictions concerning past climates based on soil magnetic mineral assemblages. For example, detailed time series of Quaternary mean annual precipitation and palaeoprecipitation gradients across wide geographic regions have been predicted from the rock magnetic properties of Chinese loess and palaeosol units. Quantitative prediction requires full assessment of the uncertainties associated with predictions. However, little attention has been given to this important aspect of climofunction prediction. We present an analysis of an ensemble of published rock magnetic climofunctions and estimate the uncertainty of the associated predictions. We find that existing climofunctions have associated uncertainties that are so large that their subsequent predictions are effectively invalid. Thus, palaeoprecipitation reconstructions must be treated with extreme caution. In the future climofunctions that are constrained geologically through the inclusion of theoretical models of soil development may provide predictions with lower uncertainties.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Numerous studies have demonstrated a relationship between climate and the magnetic properties of soils (see Orgeira et al. (2011) and Liu et al. (2012) for recent reviews). Different mechanisms, such as magnetic mineral formation during natural fires (Le Borgne, 1960; Kletetschka and Banerjee, 1995) and mixing of magnetic and non-magnetic sediments (Kukla et al., 1988; Porter et al., 2001), have been proposed to explain this relationship. It is now widely agreed, however, that the inorganic formation of secondary ferrimagnetic minerals during pedogenesis is responsible for the magnetic enhancement of soils (Maher and Taylor, 1988; Zhou et al., 1990; Maher and Thompson, 1991; Zheng et al., 1991). A link therefore exists between the climate parameters (e.g., precipitation, evaporation and ambient temperature variations) that influence soil formation and the properties of pedogenically enhanced magnetic mineral assemblages. Soil formation is, however, a complex process and the mechanisms that control the formation and destruction of magnetic minerals during pedogenesis remain a matter of debate (Orgeira et al., 2011; Liu et al., 2012).

Studies of links between magnetic properties and soil formation go beyond solely attempting to understand pedogenic processes. By quantifying the relationship between modern climate parameters and carefully selected magnetic properties of recent soils, it may be possible to develop a calibrated transfer function with which to make quantitative palaeoclimate predictions. Such magnetic “climofunctions” were pioneered by Maher et al. (1994), who considered the relationship between magnetic susceptibility and rainfall for loess–palaeosol sequences of northwest China. To isolate the portion of the magnetic mineral assemblage attributable to soil formation, Maher et al. (1994) used the pedogenic magnetic susceptibility ($\chi_{\text{ped}}$), which is simply the difference between the magnetic susceptibility ($\chi$) of pristine parent material (loess) and pedogenically enhanced subsoil. By comparing $\chi_{\text{ped}}$ of 9 modern soils from the Chinese Loess Plateau with meteorological data, Maher et al. (1994) showed that the logarithm of $\chi_{\text{ped}}$ correlates strongly with mean annual precipitation (MAP), whilst other factors, such as temperature and the time required for soil development, are of secondary importance. By assuming that this relationship does not vary with time, the derived climofunction was applied to ancient soils to make quantitative predictions of past rainfall levels and precipitation gradients across the Chinese Loess Plateau (Maher et al., 1994; Maher and Thompson, 1995; Maher and Hu, 2006).

Subsequent studies have adopted the magnetic climofunction concept of Maher et al. (1994) and have investigated the relationship between the magnetic properties of modern soils and the rainfall/temperature regime in which they developed. Some studies have used less discriminative magnetic properties, such as bulk magnetic susceptibility (Balsam et al., 2011), whilst others have developed specific parameters to quantify the abundance of fine-grained ferrimagnetic minerals that are characteristic of pedogenesis (Geiss and Zanner, 2006; Geiss et al., 2008). Irrespective of their differences in approach, the motivation behind these studies is to make quantitative predictions of past climate states based on climofunctions derived...
from the magnetic properties of modern soils. Limited geological consideration has been given to the form of the climofunctions. Instead, empirical model selection has been adopted by identifying magnetic parameters that exhibit a strong correlation with rainfall and temperature values.

Quantitative prediction of past climate states has involved determination of a regression-based calibration that links magnetic properties of modern soils to MAP or mean annual temperature (MAT) (Maher et al., 1994, 2002; Han et al., 1996; Geiss et al., 2008; Balsam et al., 2011). The calibration can then be inverted to determine a climofunction with which the magnetic properties of ancient soils provide a basis to make predictions of the environmental conditions under which a soil formed. The quality of empirical climofunctions is most often assessed using the coefficient of determination ($r^2$) between a given magnetic parameter and MAP or MAT. Given that the aim of climofunction development is to produce quantitative predictions, the $r^2$ statistic is difficult to interpret. For example, if a study quotes a predicted MAP of 600 mm/yr obtained from a climofunction with $r^2 = 0.7$, the uncertainty on the predicted MAP is not apparent in mm/yr. Some studies have attempted to quantify uncertainties associated with predictions made from empirical climofunctions. Maher and Hu (2006) and Balsam et al. (2011) presented errors in mm/yr calculated from the estimated quality of the regression-based calibration. These authors did not, however, consider additional uncertainties that arise when making predictions for samples that were not included in the original calibration data set.

The aim of this study is to demonstrate an approach with which the uncertainty associated with climofunction predictions can be quantified. Through the use of so-called discrimination intervals (Lieberman et al., 1967), the predictive power of a given climofunction can be judged. As was the case in the original cited studies, we focus on a statistical analysis that does not take into account geological constraints. The key issue is that quantification of uncertainties associated with climofunctions is essential if meaningful comparisons between predicted climate states are to be made.

2. Materials and methods

We analyse here a number of published climofunctions. These climofunctions all aim to predict MAP and are considered in sequence according to the complexity of the magnetic parameter used. As mentioned above, we do not consider the appropriateness of a given magnetic parameter to predict past rainfall levels. We focus instead on the empirical predictive power that is claimed to be associated with each climofunction in the respective cited studies.

Two data sets are taken from Balsam et al. (2011), who considered the relationship between the logarithm of $x_{\text{ped}}$ in modern soils and MAP. The first data set is composed of data from Mali ($n = 38$), whilst the second data set comprises the western subtransect ($n = 19$) subset of the Mali data.

Maher et al. 1(994) developed a climofunction for the Chinese Loess Plateau ($n = 9$) to predict MAP from the logarithm of $x_{\text{ped}}$. This work was later extended by Maher et al. (2002) to include additional soils from the Chinese Loess Plateau ($n = 31$, Porter et al. (2001)) and the Russian steppe ($n = 22$). From a statistical analysis, Maher et al. (2002) demonstrated that their models that relate $x_{\text{ped}}$ and $x_{\text{ped}}$ in Chinese and Russian soils match closely.

A study of modern loessic soils ($n = 76$) from the midwestern United States by Geiss et al. (2008) related linearly the magnetic enhancement of soil horizons (quantified by the ratio of susceptibility of anhysteretic remanent magnetisation to isothermal remanent magnetisation, $x_{\text{ARM}}$/IRM) to MAP. A statistical analysis of $x_{\text{ARM}}$/IRM revealed that it was a better predictor of regional MAP than any other previously studied magnetic parameter (Geiss et al., 2008).

Proxy calibration is assessed here for all of the selected data sets using the same form of climofunction used by the authors of the original studies (e.g., linear, log-linear). On the basis of these calibration models the power of a given magnetic parameter to predict MAP is then assessed using discrimination intervals (Lieberman et al., 1967).

2.1. Proxy calibration

Climofunction estimation is based on a process of inverse calibration, which allows predictions of an independent variable on the basis of a dependent variable (Osborne, 1991). In such calibration problems it is essential to consider if the derived climofunction will be used only once (i.e., to predict palaeoprecipitation from a single sample) or repeatedly (i.e., multiple palaeoprecipitation predictions from a collection of samples). It is reasonable to assume that after a climofunction is developed it will be used repeatedly to make numerous predictions of palaeoprecipitation at single or multiple locations as a function of time (Maher and Thompson, 1995). We therefore employ the approach of Lieberman et al. (1967) where the number of predictions to be made on the basis of an inverse calibration is considered to be arbitrarily large.

For $n$ calibration data points included in the development of a climofunction, it is assumed that the dependent parameter, $y$, can be related to the independent parameter, $x$, by the linear regression model:

$$y = a + bx + \epsilon,$$

where $\epsilon$ represents a collection of error terms. The ordinary least-squares estimator of the regression coefficients is given by:

$$b = \frac{\sum(x-x)(y-y)}{\sum(x-x)^2} \quad \text{and} \quad a = y - bx.$$

Predictions of $y$ for the independent parameter values in $x$ are then given by:

$$\hat{y} = a + bx.$$

The misfit between the data ($y$) and the regression predictions ($\hat{y}$) can be quantified by the estimated residual variance:

$$s^2 = \frac{\sum(y-y)^2}{n-2}.$$

The relationship between $x$ and $y$ obtained from the $n$ calibration data points forms a climofunction from which numerous predictions of $x$ will be made on the basis of $y$. This relationship is, however, only an estimate of the true relationship because it is based on a statistical sample of $n$ points rather than the entire population. Given this limitation, it is essential to assess the uncertainty associated with the estimated regression line. Working and Hotelling (1929) showed how a confidence band for the location of the true regression line (i.e., that of the entire population) can be determined across the range of the data. This band is based on the estimated regression line and is constructed in a point-wise manner. For example, at the point $x_0$ the confidence band is:

$$\hat{y}_s = \left\{2F_{1, n-2, \alpha/2} \right\}^{1/2} \left\{1 + \frac{(x_0-x)^2}{\sum(x-x)^2} \right\}^{1/2},$$

where $F_{1, n-2, \alpha/2}$ is the value of a $F$ distribution with $(2, n-2)$ degrees of freedom at the $1 - \alpha$ level. For a given value of $\alpha$ there is a $1-\alpha$ probability that the confidence band contains the true regression line for the population.

A second source of uncertainty originates from the ability of the parameter $y$ to make predictions of $x$. A non-zero value of $s^2$
The solutions of Eq. (13) provide the limits of the discrimination interval at \( \alpha \) at the 1 − \( \alpha \) and \( P \) level. Construction of discrimination intervals is shown graphically in Fig. 1.

2.2. Climofunction extrapolation

A worldwide study of soil \( \chi \) by Balsam et al. (2011) demonstrated an increasing trend in \( \chi \) at low MAP levels, followed by decreasing \( \chi \) at higher MAP. The turning point in the data is likely to indicate that soils have entered a saturation regime with suppressed magnetite production (Orgeira et al., 2011) or that magnetite dissolution has occurred in waterlogged soils (Han et al., 1996; Guo et al., 2001). This implies that an empirically-based climofunction can only be deemed reliable across the interval of MAP values upon which it was constructed. Balsam et al. (2011) demonstrated that a climofunction based on only low MAP values could not be reliably extrapolated to high MAP values. On the basis of this restriction the presented analysis will be limited to predicted MAP values that lie within the interval of the calibration data set.

2.3. Mapping discrimination intervals

For each data set introduced in Section 2, the value of the magnetic parameter that would yield a predicted MAP with the narrowest possible discrimination interval was found. For given values of \( \alpha \) and \( P \) the width of this discrimination interval is defined as \( \Delta \text{MAP}_{\text{min}} \) and is expressed relative to the width of the calibration interval, \( \Delta \text{MAP}_{\text{cal}} \) (Fig. 2). The ratio \( \Delta \text{MAP}_{\text{min}}/\Delta \text{MAP}_{\text{cal}} \) presents a best-case scenario (i.e., the smallest achievable discrimination interval for a given climofunction). For comparison the widest possible discrimination interval, \( \Delta \text{MAP}_{\text{max}} \), is also calculated. To assess the ability of magnetic parameters to predict palaeoprecipitation we have mapped the ratios \( \Delta \text{MAP}_{\text{min}}/\Delta \text{MAP}_{\text{cal}} \) and \( \Delta \text{MAP}_{\text{max}}/\Delta \text{MAP}_{\text{cal}} \) for a given climofunction as a function of \( \alpha \) and \( P \).

3. Results and discussion

The method presented in Section 2 provides a means with which to assess the uncertainties associated with climofunctions. Results are shown that compare the width of the discrimination intervals associated with a climofunction to the interval of the calibration data on which the climofunction is based. If a climofunction yields discrimination intervals that are narrow with respect to the span of the calibration MAP data, the resulting predictions will carry acceptable uncertainties and quantitative palaeoprecipitation reconstructions are a viable possibility. In cases where the discrimination intervals have similar spans to the calibration data, the uncertainties associated with predictions from the climofunction are large, which suggests that quantitative palaeoprecipitation reconstructions are not possible.

3.1. Mali data set

A log-linear trend that relates \( \chi \) and MAP for the Balsam et al. (2011) Mali data set yields a relatively high coefficient of determination (\( r^2 = 0.61 \)). The scatter of the data around the regression line, however, results in a wide 95% Working and Hotelling (1929) confidence band, which indicates that the predictive power of the climofunction is low (Fig. 3). The map of \( \Delta \text{MAP}_{\text{max}}/\Delta \text{MAP}_{\text{cal}} \) for the Mali climofunction represents the worst-case scenario (i.e., the largest possible discrimination interval within the calibration data set). Even in the most liberal case of \( \alpha = 0.2 \) and \( P = 0.8 \) the width of the widest discrimination interval is still ~30% larger than the span of MAP values used to derive the climofunction. For the best-case \( \Delta \text{MAP}_{\text{min}}/\Delta \text{MAP}_{\text{cal}} \) scenario, the width of the narrowest possible discrimination interval remains wider than the calibration MAP interval even at \( \alpha = 0.2 \) and \( P = 0.8 \).
Therefore, whilst the regression demonstrates that there is a significant relationship between MAP and $\chi$, the relationship effectively has no predictive power at statistically reasonable confidence levels.

3.2. Mali western subtransect

The Mali western subtransect of Balsam et al. (2011) has a coefficient of determination of 0.80 and the data have a lower level of scatter around the log-linear regression line compared to the larger Mali data set considered above (Fig. 4). This supports the conclusion of Balsam et al. (2011) that regional climofunctions can be expected to outperform those that cover larger areas. The predictive power of the western subtransect climofunction is higher than for the Mali climofunction, but it remains weak. For the most liberal case of $\alpha = 0.2$ and $P = 0.8$ both the $\Delta$MAP$_{max}/\Delta$MAP$_{cal}$ and $\Delta$MAP$_{min}/\Delta$MAP$_{cal}$ ratios are slightly below unity (Fig. 4). This implies that any discrimination intervals at these low significance levels will only have a slightly narrower span than the entire MAP interval on which the climofunction is based. For more realistic (i.e., more conservative) confidence levels of $\alpha = 0.05$ and $P = 0.95$, $\Delta$MAP$_{min}/\Delta$MAP$_{cal}$ yields

---

**Fig. 1.** Schematic illustration of the procedure for determination of a discrimination interval. (a) A linear regression model (Eq. (1)) is constructed to quantify the relationship between the independent (abscissa) and dependent variables (ordinate). (b) The 100(1 - $\alpha/2$)% Working and Hotelling (1929) confidence band for the regression line is determined from Eq. (5). (c) The 100$P$% confidence interval ($y_a \pm Q$) of the true value of the dependent variable that corresponds to a given value of the independent variable is determined (Eq. 7). (d) Determination of the discrimination interval (Eqs. (10) and (11)) for a prediction of the independent variable based on the intersections of the confidence interval on the dependent variable and the confidence band of the regression line. The width of the discrimination interval is represented as $\Delta$.}

---


Please cite this article as: Heslop, D., Roberts, A.P., Calculating uncertainties on predictions of palaeoprecipitation from the magnetic properties of soils, Global and Planetary Change (2012), http://dx.doi.org/10.1016/j.gloplacha.2012.11.013
a value of $\sim 1.6$, which demonstrates the lack of predictive power of the western subtransect climofunction.

3.3. Chinese Loess Plateau

The $\chi_{\text{ped}}$ data set of Maher et al. (1994) contains only 9 data points. Thus, whilst it has a high correlation, its predictive power can be expected to be limited (Fig. 5). All of the mapped $\Delta \text{MAP}_{\text{cal}} / \Delta \text{MAP}_m$ values are $>1$ and $\Delta \text{MAP}_{\text{min}} / \Delta \text{MAP}_{\text{cal}}$ only approaches unity for the lowest significance levels.

Maher and Thompson (1995) employed this climofunction to produce a time series of MAP estimates from the loess–palaeosol sequence at Xifeng (central Chinese Loess Plateau). The difference in MAP estimates between glacial and interglacial periods is in the 300–400 mm/yr range. The size of the discrimination intervals for the climofunction suggests, however, that even predictions of such large changes in MAP must be treated with caution. To emphasise this point, the narrowest possible discrimination interval for the climofunction at the $\alpha = 0.05$ and $P = 0.95$ level has a width of $\sim 660$ mm/yr compared to a predicted difference of 300–400 mm/yr for glacials versus interglacials.

Based on analysis of a number of loess–palaeosol sequences, Maher and Thompson (1995) produced contour maps of MAP across the Chinese Loess Plateau for different time-slices. Calculation of palaeoprecipitation contours will depend on comparison of estimates from neighbouring sites that are likely to have MAP predictions that, on the basis of discrimination interval calculations, appear to be statistically indistinguishable. Such MAP reconstructions should therefore be interpreted with caution.

3.4. Chinese and Russian soils

The combined Chinese and Russian soil data set compiled by Maher et al. (2002) has a strong correlation between $\chi_{\text{ped}}$ and MAP.
(Fig. 6). The data are, however, scattered around the log-linear regression line. As with the case of the cited Balsam et al. (2011) climofunctions, this suggests that whilst there is a significant relationship between $\chi_{ped}$ and MAP, the predictive power of the climofunction can be expected to be weak (Fig. 6).

As with the Maher et al. (1994) climofunction, $\Delta$MAP$\min$/MAP$\cal$ values only approach unity for the lowest significance levels, whilst $\Delta$MAP$\max$ is wider than the span of the MAP calibration data for all examined cases (Fig. 6). At the $\alpha=0.05$ and $P=0.95$ level, the narrowest possible discrimination interval for the climofunction is $\approx 680 \text{ mm/yr}$, compared to a $\approx 400 \text{ mm/yr}$ span of the MAP calibration data. It is important to note that the combined Chinese data sets of Maher et al. (1994) and Porter et al. (2001) span a substantially wider interval of MAP values than the data from Russian soils ($\approx 200$ and $\approx 400 \text{ mm/yr}$, respectively). Because of the narrower MAP interval spanned by the Russian soils, the predictive power for Russian sites will be lower than for the Chinese sites.

3.5. Midwestern United States

The climofunction developed by Geiss et al. (2008) for the midwestern United States and its corresponding 95% Working and Hotelling (1929) confidence band are shown in Fig. 7. As is the case for the previously discussed climofunctions, $\chi_{ARM}/\text{IRM}$ reveals a significant relationship ($r^2=0.70$) between MAP and magnetic enhancement.

In contrast to the climofunctions discussed above, the Geiss et al. (2008) relationship has higher predictive power with the unity line lying in the central regions of the $\Delta$MAP$\max$/MAP$\cal$ and $\Delta$MAP$\min$/MAP$\cal$ maps (Fig. 7). To place this in perspective, the liberal $\alpha=0.2$ and $P=0.8$ case still produces a $\Delta$MAP$\min$ of $\approx 340 \text{ mm/yr}$ and at $\alpha=0.05$ and $P=0.95$ this increases to $\approx 550 \text{ mm/yr}$ (Fig. 7). Whilst the Working and Hotelling (1929) confidence band associated with the climofunction is relatively narrow as a result of the large number of data points included in the calibration, the scatter around the regression line is large. This provides a demonstrative case, whereby the regression relationship can be defined with small uncertainties, but the data scatter yields predictions with large uncertainties (via Eq. (7)).

4. Conclusions

Using published data sets, we have assessed the ability of climofunctions to predict MAP from the magnetic properties of soils. Whilst a statistically significant correlation exists between MAP and magnetic properties, the predictive power of the climofunctions is weak. This demonstrates the fundamental difference between statistical estimation and the less certain task of statistical prediction.

The method outlined in Section 2 provides a general approach to quantifying the uncertainties associated with calibrated climofunction predictions. The studied discrimination intervals are a function of data scatter around the empirical calibration and the number of samples included in the calibration. Whilst the size of calibration data sets can be increased with additional analyses, reducing scatter in the data is more problematic. Scatter can be reduced to a certain extent by selecting magnetic properties that isolate the pedogenically enhanced component of the magnetic mineral assemblage (Geiss et al., 2008). Much of the scatter, however, is an innate property of the data that results from variations in magnetic enhancement that cannot be explained solely by MAP. This means that increasing the size of the calibration data set will reduce the discrimination intervals to a certain degree, beyond which the intrinsic scatter of the data (which is independent of data set size) becomes dominant. Future development of process-constrained climofunctions that incorporate theoretical models of pedogenically-driven magnetic enhancement (Orgeira et al., 2011) has the potential to reduce such scatter and provide predictions based on geological considerations. Regardless, to assess the extent to which palaeoprecipitation interpretations from magnetic climofunctions are quantitatively meaningful, we

Please cite this article as: Heslop, D., Roberts, A.P., Calculating uncertainties on predictions of palaeoprecipitation from the magnetic properties of soils, Global and Planetary Change (2012), http://dx.doi.org/10.1016/j.gloplacha.2012.11.013
recommend that future climofunctions are presented with a suitable analysis of their associated uncertainties, such as the type presented here. Predictions made from a climofunction should be reported with the corresponding uncertainties.

Acknowledgement

This work was supported by the Australian Research Council (grant DP110105419). The authors thank two reviewers and the Editor for constructive comments.

References


