**Climate
of the Past**

Open Access

Discussions

CREST: Climate REconstruction SofTware

M. Chevalier[1], R. Cheddadi[1], and B. M. Chase[1,2]

[1]Institut des Sciences de l'Evolution de Montpellier, UMR5554, Centre National de Recherche Scientifique/Université Montpellier 2, Bat. 22, CC061, Place Eugène Bataillon,
34095 Montpellier, cedex 5, France
[2]Department of Archaeology, History, Culture and Religion, University of Bergen,
P.O. Box 7805, 5020 Bergen, Norway

Correspondence to: M. Chevalier (manuel.chevalier@univ-montp2.fr)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

CPD

10, 625–663, 2014

CREST

M. Chevalier et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◀◀ ▶▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

Several methods currently exist to quantitatively reconstruct palaeoclimatic variables from fossil botanical data. Of these, pdf-based (probability density functions) methods have proven valuable as they can be applied to a wide range of plants assemblages. Most commonly applied to fossil pollen data, their performance, however, can be limited by the taxonomic resolution of the pollen data, as many species may belong to a given pollen-type. Consequently, the climate information associated with different species cannot sometimes not be precisely identified, resulting less accurate reconstructions. This can become particularly problematic in regions of high biodiversity. In this paper, we propose a novel pdf-based method that takes into account the different climatic requirements of each species constituting the broader pollen-type. Pdfs are fitted in two successive steps, with parametric pdfs fitted first for each species, and then a combination of those individual species pdfs into a broader single pdf to represent the pollen-type as a unit. A climate value for the pollen assemblage is estimated from the likelihood function obtained after the multiplication of the pollen-type pdfs, with each being weighted according to its pollen percentage.

To test the robustness of the method, we have applied the method to southern Africa as a regional case study, and reconstructed a suite of climatic variables based on extensive botanical data derived from herbarium collections. The reconstructions proved to be accurate for both temperature and precipitation. Predictable exceptions were areas that experience conditions at the extremes of the regional climatic spectra. Importantly, the accuracy of the reconstructed values is independent from the vegetation type where the method is applied or the number of species used.

The method used in this study is publicly available in a software package entitled CREST (Climate REconstruction SofTware) and will provide the opportunity to reconstruct reliable quantitative estimates of climatic variables even in areas with high geographical and botanical diversity.

1   Introduction

Reconstructing past climates, while being an important element in the global effort to understand climate system dynamics and their potential future structure and characteristics, is often limited to qualitative assessments of past conditions. This limits the
⁵ potential for comparisons with the general circulation model (GCM) simulations, and the integration of palaeoenvironmental information in modelling initiatives (Braconnot et al., 2012). As a result, while inconsistencies exist both between GCM simulations, and between the simulations and the fossil records, it is difficult to use the bulk of the palaeo-data available to evaluate GCM simulations in an efficient and effective way.
¹⁰ Many techniques have been developed to quantitatively reconstruct past climates from palaeo-botanical data (Guiot et al., 1993; Huntley et al., 1995; Overpeck, 1985; Kühl et al., 2002). They rely on the fundamental hypothesis that a causal relationship exists between the modern distributions of plants and the associated climates (Jackson and Williams, 2004 and references therein). These techniques can be di-
¹⁵ vided into two types: those based on plant assemblages (Modern Analogs Technique, MAT; Overpeck, 1985; Guiot, 1990, and Weighted Averaging Partial Least Square regressions, WA-PLS; ter Braak and Juggins, 1993), and those based on plant distributions (Mutual Climatic Range, MCR; Atkinson et al., 1987; Sinka and Atkinson, 1999) or probability density functions (pdfs; Kühl et al., 2002). MAT and WA-PLS are reported to
²⁰ be more accurate, but they are limited in their applicability because of the potential lack of modern analogs for some fossil pollen assemblages (Jackson and Williams, 2004). From that perspective, the flexibility of methods based on plant distribution become more advantageous to expand the range and scope of "reconstructible" environments, as they can be applied to any past assemblage providing that most species from that
²⁵ palaeo-assemblage still exist.

Conceptually, the pdf-based methods evolved from MCR techniques as a way to model the strength of the relationship between plants and climate. In fact, MCR (which considers a rectangular envelope defined by minimum and maximum values for a given

climate variable) can be seen as the most simple pdf-based method. These methods are based on the correlation between plants modern geographical distribution and climate gradients, with the climate value that is the most often found in the plant distribution being its "optimum". Among the approaches that have already been proposed within the last decade (Kühl et al., 2002; Gebhardt et al., 2007; Truc et al., 2013), a recurrent issue concerns the assumptions made about the morphological characteristics of the envelope (width, skewness, central tendencies). Kühl et al. (2002) fitted a multidimensional gaussian surface that excluded both multimodality and asymmetry that are however a common feature when dealing with botanical assemblages. Later, Gebhardt et al. (2007) proposed to fit a mixture model (combination of several gaussian surfaces) to relax the constraints of a unimodal gaussian shape. Recently, Truc et al. (2013) proposed the application of non-parametric pdfs to improve the fit between pdf and data.

In addition to the issue of the shape, the accuracy of such models is also a function of the taxonomic resolution at which pollen can be identified (usually family to generic level) and the number of species making up a given pollen-type. Pollen-types often become climatically non-informative due to a saturation effect wherein too many species result in the climatic information conveyed by each species being averaged and lost.

Contrary to the problem of the shape of the climate envelope, the problem of low taxonomic resolution has rarely been discussed as its effects are usually not significant when plant diversity is relatively low. However, in areas where pollen-types can comprise a high number of plant species (> 30), it becomes preponderant and can result in saturated pdfs. Truc et al. (2013) proposed a Species Selection Method (SSM) that recursively alter the taxonomic composition of a pollen-type by taking into account the co-existence with other pollen-types. In order to minimize pdf saturation, the SSM removes species that have climate requirements that are different from that of the assemblage.

However, the SSM only removes species with optima at the extremes of climatic gradients, leaving a certain number of climatically undifferentiated species around the

CREST

M. Chevalier et al.

◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

median climate. We believe that the problems of the shape and the diversity are in fact intimately related to the strategy used for fitting pdfs. A pollen-type is not a homogeneous ecological unit in the sense that many species with different climate requirements can be classified in the same pollen-type. From this point of view, fitting a density function directly to a pollen-type is questionable. On the basis that species are the ecological units that respond to climate gradients, we propose a two-step procedure to define the pdf of the pollen-types: (1) unimodal parametric pdfs are fitted for the species (pdf_{sp}) and (2) those parametric pdf_{sp} are combined to produce the pdf of the pollen-type (pdf_{pol}). The pdf_{pol} reflects the diversity that exists among its species by considering independently each species. To reconstruct a climate value, we propose to combine the pdf_{pol} with a weighted geometric mean. The multiplication of pdf_{pol} ensures the conservation of the mutual climatic range.

To quantify the method's capability to reconstruct different variables in different environments, we have reconstructed a set of modern climatic conditions (20 variables) over a large area (3389 quarter-degree grid cells representing southern Africa). Southern Africa – composed of South Africa, Botswana, Lesotho, Swaziland and Namibia – is well-suited for run such a test as it is characterized by a strong heterogeneity in terms of topography, geology and climate (Tyson, 1986; Partridge and Maud, 2000; Chase and Meadows, 2007) leading ultimately to a great diversity of plant species (Goldblatt and Manning, 2002). Statistical tests were performed on climate anomalies (1) to analyze where and why the model was reliable, and (2) to measure the effects of parameters such as the type of variable, the number of taxa used and/or the vegetation type.

The method presented here as been implemented in a software entitled CREST (Climate REconstruction SofTware). With its simple interface, CREST is intended to make quantitative climate reconstructions an accessible goal for the wider community. Our hope is that a proliferation of quantitative reconstructions of past climate conditions will facilitate the consideration of palaeoenvironmental data in the assessment of

CREST

M. Chevalier et al.

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

GCM performance, and ultimately allow for an improved understanding of both past and potential future climate change.

2 Methodology

The climate reconstruction method we propose is based on probability density functions (pdfs). Schematically, a pdf represents the probability of a species to exist along a climate gradient and is a surrogate for the species realised niche (see for example Kearney, 2006). The process of a pdf-based method can be divided into three successive steps: (i) quantifying the plant–climate relationship, in other words fitting the pdfs, (ii) combining information from the different taxa, and finally (iii) extracting a climate value from the resulting climate likelihood function. This method relies on a strong hypothesis: the invariance of the plant–climate relationship since the deposition of the fossil assemblage.

2.1 Fitting of the pdfs

This step is crucial for each pdf-based method. Many different strategies have been proposed (Kühl et al., 2002; Gebhardt et al., 2007; Truc et al., 2013) all of them fitting a pdf to the pollen-types identified in the fossil record. This strategy leads to a loss of certain information because (1) individual signals are mixed and (2) rare species are masked by the most extended ones.

Here we propose a two-step procedure to fit pdfs that better integrates the diversity that can exist within some pollen-types. First, we fit a pdf to each species (noted psd_{sp}), and secondly we combine the pdf_{sp} into pdf_{pol}. The latter considers more clearly the pollen-type's diversity.

2.1.1 Creating pdf$_{sp}$

Based on observations, we propose that distributions of climatic values where a species is found – its niche – can be classified into two shapes: a log-normal shape (Fig. 1a) or a normal shape (Fig. 1b) (Austin, 1987; Austin and Gaywood, 1994; Hirzel and Le Lay, 2008). The normal shape is symmetric while the log-normal shape is markedly right-skewed (left-skewed distributions have been observed but are uncommon). In addition, the log-normal function is null for negative values, which is of interest when modelling, for example, rainfall amounts. Both curves are defined by two parameters: the mean $\overline{x}_{sp}$ (Eq. 1) and the variance $s_{x,sp}^2$ (Eq. 2) of the species niche, with x being the studied climatic gradient.

Kühl et al. (2002) proposed to weight each climate observation according to its frequency of occurrence. Climate values are never equally distributed in the studied area. Consequently, to ensure homogeneity in the climate data, the rarest climate values are upweighted and the most common values are downweighted (referred later as the *climate abundance weighting*). The climate values (a total of N) are sorted into J bins of equal width. A weight k_j is defined for each bin as the ratio of N with the number of pixels n_j in the bin j.

$$\overline{x}_{sp} = \frac{1}{\sum k_i} \sum_{i=1}^{N} k_i x_i \tag{1}$$

$$s_{x,sp}^2 = \frac{1}{\sum k_i} \sum_{i=1}^{N} k_i (x_i - \overline{x}_{sp})^2 \tag{2}$$

The shape and the position of the pdf$_{sp}$ along a gradient can be calculated with Eqs. (3) and (4) representing the normal law and the log-normal law (Fig. 1b and a), respectively.

$$\mathrm{pdf}_{\mathrm{sp}}(x) \;=\; \frac{1}{\sqrt{2\pi s_{x,\mathrm{sp}}^2}}\; \exp\left(-\frac{(x-\overline{x}_{\mathrm{sp}})^2}{s_{x,\mathrm{sp}}^2}\right) \tag{3}$$

$$\mathrm{pdf}_{\mathrm{sp}}(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2 x^2}}\exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right) \;\text{with}\; \begin{cases} \mu = \ln(\overline{x}_{\mathrm{sp}}) - \frac{1}{2}\ln\left(1+\frac{s_{x,\mathrm{sp}}^2}{\overline{x}_{\mathrm{sp}}^2}\right) \\[2mm] \sigma^2 = \ln\left(1+\frac{s_{x,\mathrm{sp}}^2}{\overline{x}_{\mathrm{sp}}^2}\right) \end{cases} \tag{4}$$

2.1.2 Creating $\mathrm{pdf}_{\mathrm{pol}}$

To create the $\mathrm{pdf}_{\mathrm{pol}}$, all the $\mathrm{pdf}_{\mathrm{sp}}$ are added with a weight determined by their geographical extension (represented by the number of climate values n_i, Eq. 5). Contrary to the $\mathrm{pdf}_{\mathrm{sp}}$ that have an imposed shape, the shape of $\mathrm{pdf}_{\mathrm{pol}}$ is free, with no assumptions being made. A $\mathrm{pdf}_{\mathrm{pol}}$ can be multimodal when composed of two or more climatically separated groups of species. This is necessary to take into account the diversity within each pollen-type. Figure 1c and 1d highlights the advantage of that method: for instance, the climatic signals conveyed by the three species *Tribulus cristatus*, *T. pterophorus* and *T. zeyheri* are not masked by the signals of the most extended one *T. terrestris*.

$$\mathrm{pdf}_{\mathrm{pol}}(x) \;=\; \frac{1}{\sum \sqrt{n_{\mathrm{sp}_i}}} \sum_{\mathrm{sp}_1}^{\mathrm{sp}_N} \sqrt{n_{\mathrm{sp}_i}}\; \mathrm{pdf}_{\mathrm{sp}_i}(x) \tag{5}$$

2.2 Combination of the $\mathrm{pdf}_{\mathrm{pol}}$ to create the $\mathrm{pdf}_{\mathrm{var}}$

We propose the combination of different $\mathrm{pdf}_{\mathrm{pol}}$ with a weighted geometrical mean (Eq. 6). The multiplication of $\mathrm{pdf}_{\mathrm{pol}}$ ensures that the reconstructed climate value will be in the mutual climate range of the taxa considered. In addition, since plants pollinate more when they live close to their climate optimum (Birks and Seppä, 2004;

Jackson and Williams, 2004), the pdf_{pol} are weighted according to a monotically increasing function of their pollen percentage $\omega_{\text{pol}}(t)$ (Eq. 7).

$$\text{pdf}_{\text{var}}(x,t) = \left(\prod_{\text{pol}_1}^{\text{pol}_N} \text{pdf}_{\text{pol}_i}(x)^{\omega_{\text{pol}_i}(t)} \right)^{\left(\sum \omega_{\text{pol}_i}(t) \right)^{-1}} \tag{6}$$

Using pollen percentages $p_{\text{pol}}(t)$ to weight taxa is difficult because the pollination rate can vary substantially from one family to the other (Jackson and Williams, 2004). Kühl et al. (2002) chose not to weight the different taxa, while Truc et al. (2013) chose to rescale the percentages between 0 and 1, 1 corresponding to the highest percentage observed for the pollen-type. This strategy is, however, very sensitive to outliers. We propose instead to normalize the percentages by the mean of the percentages that are not null. For a given pollen-type, our weights have a correlation of 1 with those of Truc et al. (2013). The difference lies in the relative weights between taxa.

$$\omega_{\text{pol}}(t) = \frac{p_{\text{pol}}(t)}{\text{mean}(p_{\text{pol}}(t))_{\forall t, p_{\text{pol}}(t) > 0}} \tag{7}$$

2.3 Climate reconstruction

The reconstructed climate corresponds to the abscissa $\widehat{x(t)}$ of the optimum of pdf_{var} (Eq. 8). pdf_{var} describes the likelihood of any climatic value to be the target value when considering the presence of many pollen-types.

$$\widehat{x(t)} = \text{argmax}\left(\text{pdf}_{\text{var}}(x,t)\right) \tag{8}$$

2.4 Error estimations

pdf_{var} provide access to the complete distribution of errors. They can be estimated at different thresholds (noted α). The $\alpha\%$ confidence interval (CI) is more appropriate than a standard deviation because pdfvar are rarely symmetrical (Fig. 2).

3 CREST software

We have implemented our method into a software package entitled CREST (Climate REconstruction SofTware). CREST is an integrated multi-platform program developed to facilitate climatic reconstructions. The advantage of CREST is the opportunity to change easily a range of parameters (e.g. the shape of the pdf_{sp}, how to use the pollen percentages, using the *climate abundance weighting*, which set of pollen-types should be used, etc.). Since the optimal reconstruction of palaeoclimatic variables is an iterative process (many runs are usually necessary to interpret the reconstructed patterns) CREST can generate detailed outputs (both figures and text files) that offer the possibility to have a detailed feedback on the reconstructed values. We believe that understanding which pollen-types are important and why is of prime importance to ensure a reliable reconstruction. Many tools have been implemented to avoid the common "statistical black box" criticism and render the process accessible for the wider community.

4 Validation

As a case study, we have used a modern botanical database to reconstruct a set of contemporary climate values to highlight and explore the strengths and weaknesses of the approach, and to quantifying its accuracy and robustness. Our study area, southern Africa, is composed of five countries: South Africa, Namibia, Lesotho, Swaziland and Botswana (from 17 to 34.5° S and from 12 to 32.5° E, Fig. 3), and is composed of 3913 quarter-degree grid cells. The region is particularly suited for our experiment because the subcontinent is characterized by a strong heterogeneity in terms of topography, geology and climate (Tyson, 1986; Partridge and Maud, 2000; Chase and Meadows, 2007) leading to the existence of many vegetation types with rapid changes over short distances.

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

4.1 Climate system

Most of southern Africa is dominated by summer rainfall related to the seasonal dynamics of the Intertropical Convergence Zone (ITCZ), and the advection of moist tropical air masses off the Indian Ocean. Annual rainfall is highest along the eastern escarpment ($\sim 1200\,\text{mm}\,\text{yr}^{-1}$; Hijmans et al., 2005; Mitchell and Jones, 2005) and decreases westward. Conversely, in the Cape region (southern tip of Africa), most of the rain falls during the winter months as a result of frontal systems embedded in the southern westerlies (Tyson, 1986) and can reach a total of more than $900\,\text{mm}\,\text{yr}^{-1}$. A complex mosaic of rainfall regimes are found at the boundary between those two systems: from year-round rainfall along the south coast of South Africa ($> 900\,\text{mm}\,\text{yr}^{-1}$ distributed in more than 100 rain events per year) to the super arid Namib Desert ($< 20\,\text{mm}$, < 10 rain events). The orographic effects of the Drakensberg escarpment and the Cape Fold Belt are very marked, creating a strong rainshadow effect in their lee.

A strong west/cold to east/warm temperature gradient is also observed. The west coast is cooled by the cold northward flowing Benguela current (upwelling zone) whereas the south and east coasts are warmed by the southward flowing Agulhas and Mozambique currents, respectively. At a given latitude, the difference in temperature between the two coasts can exceed $6°\text{C}$. The greatest diurnal temperature ranges are found in the interior, especially in the Kalahari and the Karoo region where the altitude is greater than $1000\,\text{m a.s.l.}$ in many areas (Hijmans et al., 2005).

The study region currently supports four primary biomes: Deserts and Xeric Shrublands (54.7 %), Montane Grasslands and Shrublands (16.8 %), Tropical and Subtropical Grasslands, Savannas and Shrublands (25.3 %) and Mediterranean Forests, Woodlands and Scrub (3.2 %) (Olson et al., 2001). The latter is better known as the Cape Floristic Region, which is dominated by the Fynbos Biome. Each biome is divided into ecoregions (Fig. 3; Table 1), which will be used to describe the model's properties.

4.2 Data

We have extracted botanical data for all grid cells where at least one plant with more than 25 pixels in its distribution had been recorded, leading to a total of 3389 "samples" (Fig. 4). We have then selected 20 climatic variables of interest: 9 temperature-like and 5 11 moisture-like variables (Table 2). A total of 4969 species distributions have been used.

4.2.1 Botanical data

Botanical data were extracted from a series of databases held by the South African National Biodiversity Institute (SANBI, 2003; Rutherford et al., 2003, 2012). The data from 10 these sources, which are derived mainly from herbarium collections and documented observations, are available as "presence" within a particular 0.25° × 0.25° grid square. We have used this resolution for our analyses, upscaling more precisely located data to this common resolution.

In this study, we only consider species with at least 25 occurrences leading to a number of species (n_{esp}) available per pixel between 1 and 1371 (median = 47). This strong heterogeneity is mainly due to both the range of environments found in our study area (Fig. 3) and the strong difference that exists between the different countries (Fig. 4), with South Africa providing by far the most extensive dataset.

4.2.2 Climatic data

To define pdfs, the species distributions have to be associated with climate data. We have used a subset of the climatic variables from WORLDCLIM1.4 (Hijmans et al., 2005), which, along with monthly precipitation and temperature data, provides a dataset of 19 bioclimatic variables that are considered important elements in studying the eco-physiological tolerance of plants species. These data were then upscaled 25 to match the resolution of the botanical data (0.25° × 0.25°, quarter-degree). Additional

variables of interest have also been derived from WORLDCLIM's data, including the soil water content (SWC; Trabucco and Zomer, 2010) for both summer and winter, the mean annual aridity (Trabucco and Zomer, 2009) and the percentage of winter rainfall (WRP). WRP and WORLDCLIM's Precipitation Seasonality variable are similar as they
5 measure the distribution of rainfall across the year. Summer and winter rainfalls are not differentiated by WORLDCLIM's Precipitation Seasonality, however. As this distinction is important for understanding past climate changes over southern Africa, we have created WRP to try to consider this major difference in our reconstructions.

We have also used two variables from the CRU 2.10 time series (Mitchell and Jones,
10 2005): the number of frost and wet days during the year. Those data ($0.5° \times 0.5°$ grid cells) were downscaled to meet our resolution.

The description of all variables as well as their original reference is summarized in Table 2.

5 Results

15 ### 5.1 Accuracy of the model

We have measured the climate anomalies $\delta(v, s)$ for each sample s and each variable v between the reconstructed climate Recon(v, s) and the instrumental value Instru(v, s) according to Eq. (9). A positive/negative anomaly is equivalent to an under/over-estimation of the targeted climate.

20 $$\delta(v, s) = \text{Instru}(v, s) - \text{Recon}(v, s) \tag{9}$$

The dispersion of the anomalies for each variable has been compiled in Table 3. The distributions of anomalies are all centered around 0, meaning that the model is not subject to undue bias. Nevertheless, a major dichotomy can be observed between
25 the two types of variables: for the temperature-like variables, the median is positive for eight out of nine variables (general under-estimation) while for moisture-like variables

Printer-friendly Version

Interactive Discussion

the opposite is observed, with negative medians for eight out of eleven variables (general over-estimation). The different percentiles we have calculated give insight about the dispersion of the reconstructed values as well as do the histograms in Fig. 9. The skewness is most often negative (for 14 variables), meaning that when errors are negative (over-estimation) their absolute value is higher than when they are positive (75 and 95 % percentiles respectively higher than the 25 and 5 %).

The Root Mean Square Deviation (RMSD(v); Eq. 10) is an index that reflects the mean error of a model but it has the inconvenient to be sensitive to outliers. It allows, however, for a good evaluation of the performance of the model. All the values are compiled in Table 3.

$$\text{RMSD}(v) \; = \; \sqrt{\frac{1}{N} \sum_{s=1}^{N} \delta(v,s)^2} \qquad\qquad (10)$$

The amplitude of $\delta(v,s)$ and RMSD(v) are functions of the variable range. Direct comparisons between variables cannot be performed – except for those with a similar range of variation such as, for example, Tmean Ann, Tmean Cold Q and Tmean Warm Q. To remove this discrepancy, we have normalized our RMSDs by the observed standard deviation of the instrumental values (NRMSD(v); Eq. 11). NRMSDs are lower for moisture like-variables whilst they exhibit the highest anomalies (in units of NRMSD; Figs. 6 and 7). Four variables present a high NRMSD: Mean Diurnal Range (0.75), Temp Ann Range (0.72), Prec seasonality (0.70) or Temp seasonality (0.68). The climatic signal of these four variables does not seem to be well captured by the botanical data, and plant distribution are apparently not directly driven by those variables. They represent annual climatic variability, and a range of climatic scenarios could result in the same values. For example, the variable Prec Seasonality takes identical values for seasonal rainfalls whether they occur mainly in winter or summer. This major difference is incorporated in WRP that has been reconstructed with a much better accuracy (NRMSD = 0.44).

$$\text{NRMSD}(v) \;=\; \frac{\text{RMSD}(v)}{\sigma_{\text{Instru}(v)}} \tag{11}$$

5.2 Geographical analysis of the errors

Generally, southern African climates are accurately reconstructed by CREST. The anomalies that do exist are not randomly dispersed throughout the study area. On the contrary, regions of enhanced or diminished error are observed for each variable (Figs. 5 and 6). On these figures the anomaly has been normalized by the RMSD (Eq. 12) to make all the maps comparable. In addition, only the absolute value is considered. This observation is validated with the measure of the spatial autocorrelation of the anomalies with Moran's I (Moran, 1950) (Table 3). To compute this index a neighborhood matrix of weights has to be defined. We have considered that the neighbors of a grid cell are only the 8 grid cells directly adjacent to it. All the values are between 0.31 (Prec Ann) and 0.70 (Prec Seasonality) while under the null hypothesis (no spatial autocorrelation) the expected value for Moran's I is -3.95×10^{-4} (variance of 8.25×10^{-5}). Moran's I is normally distributed, so that our results demonstrate that the anomalies are spatially clustered: in some areas the model performed very well while it failed in some others.

$$\delta_{\text{norm}}(v, s) \;=\; \left| \frac{\delta(v, s)}{\text{RMSD}(v)} \right| \tag{12}$$

Four areas present a group of outliers for several variables: (1) the Namibian coast (for temperatures and precipitations), (2) the high mountains of Lesotho (for temperatures), (3) the Eastern part of the Great Escarpment (precipitations) and (4) the southern coast of South Africa (precipitations) (Figs. 5 and 6).

5.3 Factors impacting the reconstructions

Errors being spatially clustered, we have looked for factors that could explain this concentration. There is no clear linear relation between the anomalies absolute values and n_{esp}. The slopes of the linear models we fitted were statistically significant at the 5 % threshold but the R^2 were always low (3.3 % of variance explained in average, Table 4). The models can, however, be biased by the uneven distribution of n_{esp}; half of the grid cells were reconstructed with 47 or less species while some others were reconstructed with more than 1000 (Fig. 4). Some of our clusters of errors are found in mountainous regions, and we have hypothesized that the errors may arise from a mix of low altitude plants with high altitude plants, with the anomalies observed being proportional to the degree of mixing. Thus, we have calculated the intra-pixel variation of altitude (the standard deviation of all the 30 arc-second altitude values in each quarter-degree grid cell, later called ΔAlt). We fitted linear model to explain the anomalies as a function of n_{esp} and ΔAlt. However, the gain of explained variance was relatively small (+0.9 % in average). These results indicate that the anomalies are not a result of the number of species used.

We also considered the impact of vegetation type on the anomalies. We used Olson et al. (2001) classification to assign a biome and an ecoregion to each grid cell (Fig. 3). We used an ordination technique called Between-Groups PCAs (Thioulouse et al., 1997) to reveal the differences that may exist between vegetation types. With all the variables considered in the same analysis, we measured if the type of vegetation impacted the reconstructions. At the biome level (7 levels; Fig. 7), the between-groups variance only explained 9 % of the total variance, meaning that more than 90 % of the variance was not explained by the differences between the biomes. The length of the boxes on Fig. 7 highlights that there is more variance within each group than between them.

The Between-Groups PCA ran on the ecoregions explains 25 % of the total variance but this is low relative to the number of levels (25). Again, more variance remained

within the groups than between them. Figure 8 summarises the mean dispersion of errors within each ecoregion. Some ecoregions appears to concentrate outliers, but these are always composed of 25 or less samples (low geographical extension and/or low amount of botanical data). Thus, despite the high diversity that exists in southern
5 Africa, we were not able to demonstrate that the type of vegetation (forests, grasslands, savannas, etc.) had any effect on the quality of the reconstructions.

Finally, the only factor that explains a significant part of the dispersion is the distance of the targeted value from the most represented value of the variable over the study area (Table 4). We fitted linear models to explain the anomalies as a function of the
10 targeted value (Fig. 9). All were significant ($p_{value} < 0.001$) with positive slopes. A noticeable difference between temperature-like ($R^2 = 42\%$ in average) and moisture-like variables ($R^2 = 22\%$ in average) is observed. In others words, these results mean that values that lie far from the most represented climate exhibit the highest anomalies (on the left and/or right-hand side(s) on the x axes in Fig. 9). As three different phases
15 were identifiable in the dispersion of the anomalies along their climate gradient in most cases, we also tried to fit a linear model with third-order polynomial of the targeted climate but the increase in explained variance was not sufficient to accept this model (+4.9 % in average).

6 Discussion

20 Our results indicate that the pdf-based method employed by CREST is robust (Table 3, Figs. 5 and 6), even if some differences exist between variables. The variables that were best reconstructed were those that have a direct impact on the physiology of plants, and thus strongly constrain their distribution (e.g. TmeanWetQ, Frost Days, PrecDryQ or PrecWetQ) (referred to as *direct gradients* by Guisan and Zimmermann,
25 2000). The impact of other variables such as Mean Diurnal Range or Temp Seasonality is less direct, leading to a loss of reproducibility. In the semi-arid to arid environments of southern Africa, precipitation and/or water availability strongly constrain

plants distributions. It is thus not surprising to get lower NRMSDs for moisture-related variables. This distinction between direct and indirect gradients is crucial when performing long-term climate reconstructions. In general, we can conclude that variables that influence directly plant distributions will have a greater chance of being accurately reconstructed.

The method performed well regardless of vegetation type. We were not able to show any differences in accuracy between the different biomes and/or ecoregions providing that the distribution of the biome and/or ecoregion was sufficiently spatially extensive. Based on these results, we have found that the method works best for vegetation types represented by at least ∼ 25 to 50 quarter-degree grid cells (estimation based on Fig. 8) in order to adequately determine the plant–climate relationship.

While our expectation was that a high number of species would result in more precise reconstructions, we were not able to observe any relationship between anomalies and the number of species. Anomalies do decrease when the number of species begins to increase (from 1 to ∼ 20–30), but then the tendency is reversed, and the large anomalies were observed in samples with the largest number of species. This may be related to a saturation problem, wherein more is not necessarily better. As we used a presence/absence weighting strategy, species far from their climate optimum have the same importance as those living in their optimal climate. The increase in the number of species could increase these marginal elements, biasing the reconstructions. The role of the number of taxa on the accuracy is not yet fully understood and is the subject of ongoing studies.

Other studies (Kühl et al., 2002; Truc et al., 2013; Scott et al., 2003) have shown that selecting a subset of the recorded taxa was sometimes more appropriate when attempting to capture a given climate signal. In order to improve the quality of the reconstructed variables, consideration should be given to reconstructing each variable with a different subset of the total of the available species list. Reducing this list to a shorter list of responsive species reduces the noise and consequently leads to better reconstructions. These choices are always subjective, however, and should be done

CREST

M. Chevalier et al.

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

through a consideration of the ecology of the given species (and its known or inferred sensitivity to certain climate variables), and/or an examination of the derived pdfs (flat and multimodal pdfs may indicate insensitivity to a given climate variable). The process of selection is thus not straightforward, and while it may improve reconstructions, care needs to be taken to avoid undue bias in the results. CREST provides a range of outputs that indicate the sensitivity of different taxa to given climatic parameters, and allow the user to assess the data being considered, and make informed choices in the selection of such subsets.

When plotted on a map (Figs. 5 and 6), the anomalies appear to be spatially clustered. Those patches of large anomalies can be explained by the position of the local climate along the climate gradients (Fig. 9) and are a direct consequence of the hypotheses underlying the model. The method is correlative, and consequently it is biased towards the best represented climate values. In most cases, lowest/highest values along the studied climate gradient have few occurrences even if there are exceptions. For example, low rainfall amounts are common in southern Africa, and as a result they are well represented and the signal easily captured by the model.

To offset the impact of the climate distribution's heterogeneity, we upweighted rare climate values as proposed by Kühl et al. (2002) and Truc et al. (2013). This method shifts pdfs optima towards the rarest climate values. The climate abundance weighting did decrease the errors for the extreme climates but also increased them for the most common. The global impact is nevertheless positive since it decreased the RMSDs for all the variables. It also reduced the clustering of errors. Despite its advantages, the strategy has one major drawback in that artificial geographical limits must be selected (e.g. mountain ranges or country borders; Kühl et al., 2002) to compute the weights. A finite number of grid cells must be selected and sorted into bins. Any change in the boundaries would affect – potentially drastically – the weights, and thus the reconstructions.

Even with the climate abundance weighting, it is apparent that reconstructing the rarest climates is extremely complicated with models such as those described here.

This is why, for example, the Cape region is poorly reconstructed for the precipitation-like variables but not for the temperature-like variables. The temperature of the area is common in southern Africa – so its signal is well captured – but it is an outlier in terms of quantity and seasonality of rainfall. Other areas of notable climatic rarity in south-
5 ern Africa include: (1) the eastern portion of the Great Escarpment (high precipitation), (2) the high mountains of Lesotho where temperatures are very low and precipitation is high, (3) the thin coastal band along the southern coast of South Africa where moist forests can develop as a result of significant aseasonal rainfall, and (4) along the Namibian coast (stable temperature and extremely low precipitation). All these areas
10 lie at an extreme of one or several climatic gradients, giving raise to clusters of high anomalies. It should be stated that the notion of "extreme" is relative to the study area.

The value of reconstructing quantitatively long-term climate variations from fossil bio-logical proxies is evident. Nonetheless, the limitations of the statistical methods applied should be considered. Salonen et al. (2013) have shown – using a weighted-averaging
15 (WA) regression – that the accuracy of quantitative reconstructions based on fossil biological proxies relied strongly on the calibration dataset. By selecting randomly gen-erated calibration datasets, they were able to show that for a given site their method (1) produced different climate reconstructions in terms of values and/or amplitudes for each calibration dataset but that (2) the reconstructed patterns were all very similar.
20 We believe that a similar effect could be acting in our model with pdf_{sp} being biased by the modern climatic space. Climatic space varies over time, and some elements of certain past climate regimes may not be found and/or accessible to some species in the modern climatic space (Veloz et al., 2012). Depending on the location of the site vis-à-vis the climatic space, the potential to estimate the amplitude of climate change
25 varies. As shown schematically on Fig. 10, samples located in the mean climate space have greater potential to "move" in several directions and with greater amplitude than samples that are already at the margin of the climatic space. In the latter case, the exact amplitude of change may be underestimated, but the overall trends and direction of change may still be accurate. It is expected that even under a different climate, the

CREST

M. Chevalier et al.

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

relative position of the different taxa along a climatic gradient would stay the same, so that the replacement in the past of a taxon by another that currently lives in colder environments will effectively indicate colder conditions with – possibly large – uncertainties regarding the amplitude of change (Veloz et al., 2012).

7 Conclusions

The pdf-based method we have presented in this paper provides robust results across a range of climates and vegetation types. We have demonstrated that the accuracy does not vary significantly as a function of vegetation type or the number of species considered, and it is thus a useful tool for reconstructing climates in many regions and biomes. The accuracy of the reconstructions is, however, strongly impacted by the climate variable being reconstructed (direct or indirect gradients) and primarily by the position of the targeted climate on the climate gradient of the study area. To ensure a robust reconstruction, one should:

1. select climate variables that directly impact the distribution of the species, and, inversely, use only species whose distributions are significantly defined by the climatic variable;

2. work with samples collected in widespread vegetation types to fit the most reliable pdfs;

3. define a climatically coherent study area to take advantage of the *climate abundance weighting*.

The results presented in this paper highlight our current understanding of the potential and limitations of CREST for reconstructing climates from botanical data. Recent work has shown the potential of the models upon which CREST has been based, particularly in regards to long-term climate reconstructions (Truc et al., 2013). Our goal with CREST is to make these techniques more accessible to the wider scientific community

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

and it is our hope that this tool will be applied to study other areas where long-term climate variations still need to be quantitatively described.

CREST is freely accessible on simple demand to the authors and/or on www.(domain as yet undefined).com.

References

Atkinson, T. C., Briffa, K. R., and Coope, G. R.: Seasonal temperatures in Britain during the past 22,000 yr, reconstructed using beetle remains, Nature, 325, 587–592, 1987. 627

Austin, M. P.: Models for the analysis of species' response to environmental gradients, Vegetatio, 69, 35–45, 1987. 631

Austin, M. P. and Gaywood, M. J.: Current problems of environmental gradients and species response curves in relation to continuum theory, J. Veg. Sci., 5, 473–482, 1994. 631

Birks, H. J. B. and Seppä, H.: Pollen-based reconstructions of late-Quaternary climate in europe – progress, problems, and pitfalls, Acta Palaeobot., 44, 317–334, 2004. 632

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nat. Clim. Change, 2, 417–424, 2012. 627

Chase, B. M. and Meadows, M. E.: Late Quaternary dynamics of southern Africa's winter rainfall zone, Earth-Sci. Rev., 84, 103–138, 2007. 629, 634, 656

Deacon, J. and Lancaster, N.: Late Quaternary palaeoenvironments of southern Africa, Clarendon Press, Oxford, 225 pp., 1988.

Gebhardt, C., Kühl, N., Hense, A., and Litt, T., Reconstruction of quaternary temperature fields by dynamically consistent smoothing, Clim. Dynam., 30, 421–437, 2007. 628, 630

Goldblatt, P. and Manning, J. C.: Plant diversity of the Cape region of southern Africa, 2002, Ann. Mo. Bot. Gard., 89, 281–302, 2002. 629

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Guiot, J.: Methodology of the last climatic cycle reconstruction in France from pollen data, Palaeogeogr. Palaeocl., 80, 49–69, 1990. 627

Guiot, J., de Beaulieu, J. L., Cheddadi, R., David, F., Ponel, P., and Reille, M.: The climate in Western Europe during the last GlacialInterglacial cycle derived from pollen and insect remains, Palaeogeogr. Palaeocl., 103, 73–93, 1993. 627

Guisan, A. and Zimmermann, N. E.: Predictive habitat distribution models in ecology, Ecol. Model., 135, 147–186, 2000. 641

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, 2005. 635, 636, 651

Hirzel, A. H. and Le Lay, G.: Habitat suitability modelling and niche theory, J. Appl. Ecol., 45, 1372–1381, 2008. 631

Huntley, B., Berry, P. M., Cramer, W., and McDonald, A. P.: Modelling present and potential future ranges of some European higher plants using climate response surfaces, J. Biogeogr., 22, 967–1001, 1995. 627

Jackson, S. T. and Williams, J. W.: Modern analogs in Quaternary Paleoecology: Here today, gone yesterday, gone tomorrow, Annu. Rev. Earth Pl. Sc., 32, 495–537, 2004. 627, 633

Kearney, M.: Habitat, environment and niche what are we modelling, OIKOS, 115, 186–191, 2006. 630

Kühl, N., Gebhardt, C., Litt, T., and Hense, A.: Probability density functions as botanical-climatological transfer functions for climate reconstruction, Quaternary Res., 58, 381–392, 2002. 627, 628, 630, 631, 633, 642, 643

Mitchell, T. D. and Jones, P. D.: An improved method of constructing a database of monthly climate observations and associated high-resolution grids, Int. J. Climatol., 25, 693–712, 2005. 635, 637, 651

Moran, P. A. P.: Notes on continuous stochastic phenomena, Biometrika, 37, 17–23, 1950. 639

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., and Kent, J.: Biodiversity hotspots for conservation priorities, Nature, 403, 853–858, 2000.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial ecoregions of the world a new map of life on Earth, BioScience, 51, 933–938, 2001. 635, 640, 650, 656

CREST

M. Chevalier et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Overpeck, J. T.: A pollen study of a late Quaternary peat bog, south-central Adirondack Mountains, New York, Geol. Soc. Am. Bull., 96, 145–154, 1985. 627

Partridge, T. C. and Maud, R. R.: The Cenozoic of Southern Africa, Oxford University Press, Oxford, 2000. 629, 634

5 Rutherford, M. C., Powrie, L. W., and Midgley, G. F.: ACKDAT a digital spatial database of distributions of South African plant species and species assemblages, S. Afr. J. Bot., 69, 1–6, 2003. 636

Rutherford, M. C., Mucina, L., and Powrie, L. W.: The South African National Vegetation Database History, development, applications, problems and future, S. Afr. J. Sci., 108, 629, doi:10.4102/sajs.v108i1/2.629, 2012. 636

10 Salonen, J. S., Seppa, H., and Birks, H. J. B.: The effect of calibration data set selection on quantitative palaeoclimatic reconstructions, Holocene, 23, 1650–1654, 2013. 644

PRECIS (National Herbarium Pretoria (PRE) Computerized Information System): database, available at: http://sibis.sanbi.org/faces/DataSources.jsp (last access: 28 December 2011),

15 2003. 636

Scott, L., Holmgren, K., and Talma, A.: Age interpretation of the Wonderkrater spring sediments and vegetation change in the Savanna Biome, Limpopo province, South Africa, S. Afr. J. Sci., 99, 484–488, 2003. 642

Sinka, K. J. and Atkinson, T. C.: A mutual climatic range method for reconstructing palaeocli-

20 mate from plant remains, J. Geol. Soc. Lond., 156, 381–396, 1999. 627

ter Braak, C. J. F., and Juggins, S., Weighted averaging partial least squares regression (WA-PLS) an improved method for reconstructing environmental variables from species assemblages, Hydrobiologia, 269–270, 485–502, 1993. 627

Thioulouse, J., Chessel, D., Doledec, S., and Olivier, J. M.: ADE-4 a multivariate analysis and

25 graphical display software, Stat. Comput., 7, 75–83, 1997. 640

Trabucco, A. and Zomer, R. J.: Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database, CGIAR Consortium for Spatial Information, Published online, available from the CGIAR-CSI GeoPortal at: http://www.csi.cgiar.org (last access: January 2013), 2009. 637, 651

30 Trabucco, A., and Zomer, R. J.: Global Soil Water Balance Geospatial Database, CGIAR Consortium for Spatial Information, Published online, available from the CGIAR-CSI GeoPortal at: http://www.cgiar-csi.org (last access: January 2013), 2010. 637, 651

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Truc, L., Chevalier, M., Favier, C., Cheddadi, R., Meadows, M. E., Scott, L., Carr, A. S., Smith, G. F., and Chase, B. M., Quantification of climate change for the last 20,000 yr from Wonderkrater, South Africa: implications for the long-term dynamics of the Intertropical Convergence Zone, Palaeogeogr. Palaeocl., 286, 575–587, 2013. 628, 630, 633, 642, 643, 645, 654

Tyson, P. D.: Climatic Change and Variability in Southern Africa, Oxford University Press, Cape Town, 208 pp., 1986. 629, 634, 635

Veloz, S. D., Williams, J. W., Blois, J. L., He, F., Otto-Bliesner, B., and Liu, Z.: No-analog climates and shifting realized niches during the late quaternary: implications for 21st-century predictions by species distribution models, Global Change Biol., 18, 1698–1713, 2012. 644, 645

CREST

M. Chevalier et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Description of the seven southern African biomes, with their respective ecoregions (Olson et al., 2001), the countries in which they are found, and the number of grid cells reconstructed for each ecoregion.

Biome	Ecoregion	Countries	Grid Cells
Deserts and Xeric Shrublands (yellow)	Kalahari xeric savanna	Botswana, Namibia, South Africa	656
	Kaokoveld desert	Namibia	25
	Nama Karoo	Namibia, South Africa	513
	Namib desert	Namibia	75
	Namibian savanna woodlands	Namibia	240
	Succulent Karoo	Namibia, South Africa	151
Flooded Grasslands and Savannas (blue)	Etosha Pan halophytics	Namibia	10
	Zambezian flooded grasslands	Botswana, Namibia	35
	Zambezian halophytics	Botswana	25
Mangroves (white)	Southern Africa mangroves	South Africa	1
Mediterranean Forests, Woodlands and Scrub (purple)	Albany thickets	South Africa	29
	Lowland fynbos and renosterveld	South Africa	53
	Montane fynbos and renosterveld	South Africa	75
Montane Grasslands and Shrublands (brown)	Highveld grasslands	Lesotho, South Africa	272
	Drakensberg alti-montane grasslands and woodlands	Lesotho, South Africa	19
	Drakensberg montane grasslands, woodlands and forests	Lesotho, South Africa, Swaziland	306
	Maputaland–Pondoland bushland and thickets	South Africa	29
Tropical and Subtropical Grasslands, Savannas and Shrublands (green)	Angolan Mopane woodlands	Namibia	101
	Kalahari Acacia–Baikiaea woodlands	Botswana, Namibia, South Africa	300
	Southern Africa bushveld	South Africa, Botswana	204
	Zambezian and Mopane woodlands	Botswana, Swaziland, South Africa, Namibia	128
	Zambezian Baikiaea woodlands	Botswana, Namibia	99
Tropical and Subtropical Moist Broadleaf Forests (grey)	Knysna–Amatole montane forests	South Africa	1
	KwaZulu–Cape coastal forest mosaic	South Africa	20
	Maputaland coastal forest mosaic	South Africa, Swaziland	22

Table 2. List of the 20 climate variables reconstructed for southern Africa (name, description and original reference).

	Variable's name	Description	Reference
Temperature	Tmean ann	Mean annual temperature	Hijmans et al. (2005)
	Mean Diurnal Range	Mean of monthly (max temp – min temp)	Hijmans et al. (2005)
	Temp seasonality	Standard deviation of the annual temperature (*100)	Hijmans et al. (2005)
	Temp ann range	Annual range of temperature (max – min)	Hijmans et al. (2005)
	Tmean Wet Q	Mean temperature of the wettest quarter	Hijmans et al. (2005)
	Tmean Dry Q	Mean temperature of the driest quarter	Hijmans et al. (2005)
	Tmean Warm Q	Mean temperature of the warmest quarter	Hijmans et al. (2005)
	Tmean Cold Q	Mean temperatures of the coldest quarter	Hijmans et al. (2005)
	Frost days	Number of frost days per year	Mitchell and Jones (2005)
Moisture	Prec ann	Annual precipitations	Hijmans et al. (2005)
	Prec seasonality	Coefficient of variation of annual precipitations	Hijmans et al. (2005)
	Prec Wet Q	Precipitations of the wettest quarter	Hijmans et al. (2005)
	Prec Dry Q	Precipitations of the driest quarter	Hijmans et al. (2005)
	Prec Warm Q	Precipitations of the warmest quarter	Hijmans et al. (2005)
	Prec Cold Q	Precipitations of the coldest quarter	Hijmans et al. (2005)
	SWC Winter	Soil water content during winter	Trabucco and Zomer (2010)
	SWC Summer	Soil water content during summer	Trabucco and Zomer (2010)
	Aridity	Mean annual aridity index (*1000)	Trabucco and Zomer (2009)
	WRP	Percentage of winter rainfall	derived from Hijmans et al. (2005)
	Wet days	Number of rain days per year	Mitchell and Jones (2005)

CREST

M. Chevalier et al.

Table 3. Summary of the dispersion of the anomalies of each variable.

		5 %	25 %	50 %	75 %	95 %	Skewness	RMSD	NRMSD	Moran's I
Temperature	Tmean ann	−1.87	−0.22	0.62	1.37	2.65	−0.111	1.51	0.56	0.551
	Mean Diurnal Range	−2.63	−0.15	0.51	1.09	2.16	−1.568	1.58	0.75	0.663
	Temp seasonality	−1324.63	−294.97	77.49	345.97	765.66	−1.206	662.29	0.68	0.653
	Temp ann range	−6.07	−0.83	0.7	1.9	4.06	−1.42	3.25	0.72	0.639
	Tmean Wet Q	−2.23	−0.29	0.71	1.61	3.36	0.455	1.99	0.56	0.524
	Tmean Dry Q	−3.33	−0.76	0.24	1.33	3.02	−0.552	2.03	0.55	0.505
	Tmean Warm Q	−2.74	−0.34	0.61	1.44	2.63	−0.576	1.69	0.65	0.635
	Tmean Cold Q	−1.75	−0.22	0.6	1.51	3.24	0.392	1.7	0.55	0.528
	Frost days	−22.16	−9.37	−3.43	2.7	16.32	−0.157	12.7	0.52	0.501
Moisture	Prec ann	−243.67	−96.34	−39.21	0.14	81.11	−1.711	125.67	0.56	0.31
	Prec seasonality	−10.96	−0.1	5.8	17.89	36.02	1.02	17.83	0.7	0.696
	Prec Wet Q	−102.53	−41.41	−14.3	9.26	59.73	−0.505	56.33	0.52	0.42
	Prec Dry Q	−24.01	−10.51	−5.38	−0.27	13.19	1.046	14.42	0.57	0.566
	Prec Warm Q	−143.6	−54.28	−18.97	0.66	41.49	−1.13	68.3	0.63	0.469
	Prec Cold Q	−26.65	−9.27	−4.24	−0.19	12.55	0.954	19.51	0.48	0.447
	SWC Winter	−53.89	−22.03	−11.89	−2.51	15.5	−0.975	26.87	0.63	0.481
	SWC Summer	−49.24	−22.47	−9.82	−1.52	17.97	−0.89	25.77	0.51	0.323
	Aridity	−1718.68	−740.92	−354.43	−62.4	687.15	−0.592	929.24	0.53	0.341
	WRP	−12.84	−5.1	−1.75	0.47	4.46	−2.055	7.31	0.44	0.332
	Wet days	−22.31	−10.69	−5.68	0.13	10.74	−0.883	12.16	0.5	0.437

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 4. Percentages of variance (R^2) explained for the different factors we tested in this study.

	n_{esp}	n_{esp}*ΔAlt	biomes	ecoregions	clim	poly(clim,3)
Tmean ann	2.13	14.44	14.68	24.19	48.76	49.85
Mean Diurnal Range	1.01	11.34	6.36	41.17	60.00	63.96
Temp seasonality	1.29	6.83	5.80	42.78	39.73	51.36
Temp ann range	0.02	9.62	8.40	45.13	43.89	53.06
Tmean Wet Q	0.77	9.02	4.57	20.25	26.74	28.90
Tmean Dry Q	1.95	6.35	11.01	21.63	29.01	29.42
Tmean Warm Q	0.45	18.91	9.79	33.28	50.81	50.90
Tmean Cold Q	1.63	4.00	7.97	23.86	44.56	46.06
Frost days	1.96	2.42	4.22	15.18	37.17	38.95
Prec ann	6.19	6.26	6.07	14.38	5.14	10.21
Prec seasonality	13.44	15.69	17.74	45.77	59.48	63.43
Prec Wet Q	1.24	5.47	1.80	15.74	9.42	13.74
Prec Dry Q	8.49	9.58	18.70	28.79	49.33	53.42
Prec Warm Q	4.03	4.10	10.98	20.67	4.69	12.22
Prec Cold Q	6.82	7.65	10.42	16.40	28.83	38.39
SWC Winter	1.19	1.65	14.48	24.41	20.39	23.82
SWC Summer	2.98	3.74	3.06	13.34	8.05	18.50
Aridity	4.68	5.62	7.99	16.76	18.31	28.85
WRP	4.18	4.52	6.51	16.63	8.23	12.06
Wet days	2.09	2.92	10.43	19.51	24.65	28.39
Mean	3.33	7.51	9.05	24.99	30.86	35.77
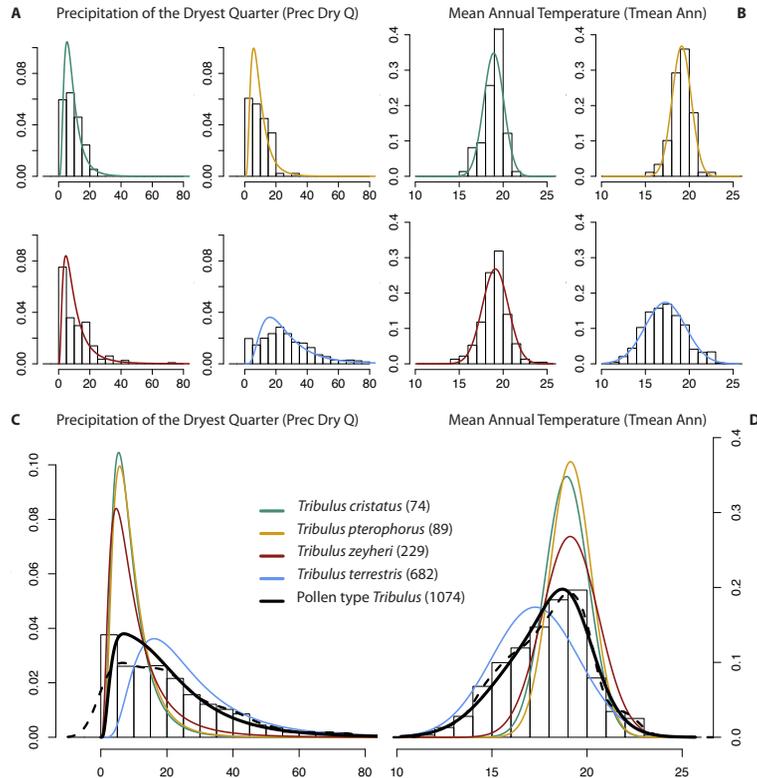
Fig. 1. Fitting of the pdfs is exemplified for two variables (Prec Dry Q and Tmean Ann) and the pollen type *Tribulus*, which is composed of four species in our database. Four pdf_{sp} are then fitted for each variable (**A** and **B**), and combined to create the pdf_{pol} (**C** and **D**). The dashed lines on **(C)** and **(D)** are the pdfs obtained by Truc et al. (2013). The difference between the two methods is more marked for Prec Dry Q where (i) the pdf_{pol} is null for negative precipitation values (more realistic) and (ii) the optimum is more marked and reflects the optima of the different species.
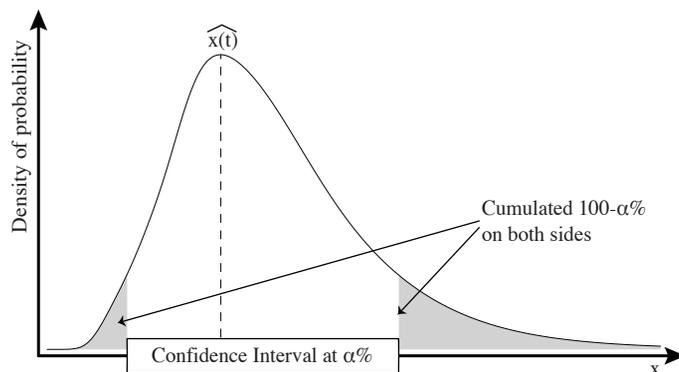
Fig. 2. Calculation of a CI exemplified with a right-skewed pdf_{var}. More values are rejected on the right-hand side of the climate gradient. The grey areas cover an area representing $\alpha\%$.

Full Screen / Esc
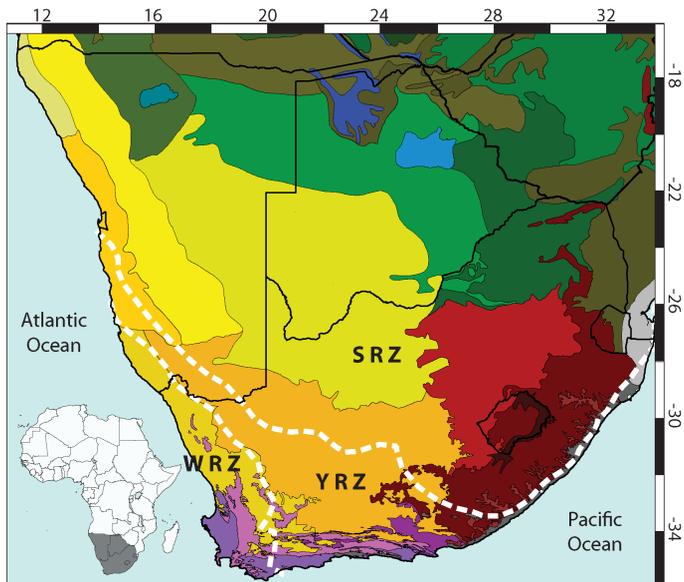
Fig. 3. Distribution of the different biomes and ecoregions in southern Africa (Olson et al., 2001). "Mediterranean Forests, Woodlands and Scrub" is in purple, "Deserts and Xeric Shrublands" in yellow, "Montane Grasslands and Shrublands" in brown, "Tropical and Subtropical Grasslands, Savannas and Shrublands" in green, "Flooded Grasslands and Savannas" in blue and "Tropical and Subtropical Moist Broadleaf Forests" in grey. Mangroves also exist along the southeastern coast of South Africa and are represented in white. The dashed white lines delineates the different rainfall zones as defined by Chase and Meadows (2007): the Winter Rainfall Zone (WRZ; > 66 % winter rain), the Summer Rainfall Zone (SRZ; < 33 % of winter rain) and the Year-round Rainfall Zone (YRZ) in between.

Full Screen / Esc
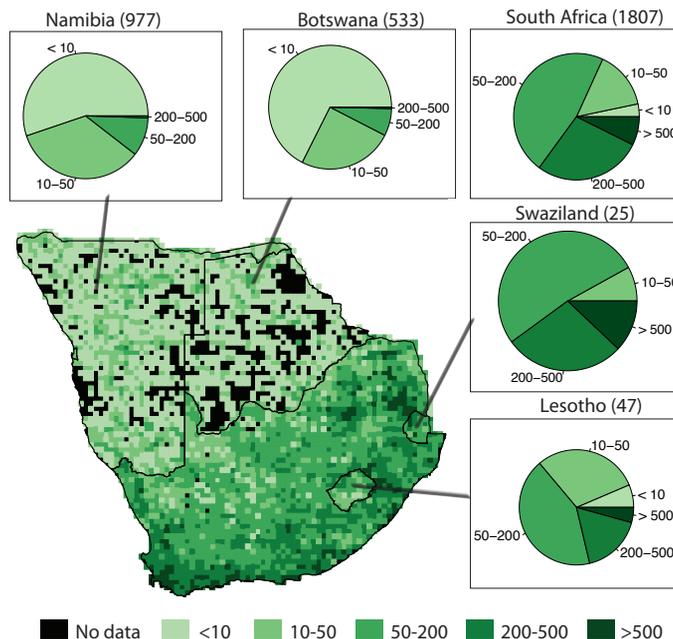
Fig. 4. Distribution of the number of species per grid cell. The greener the grid cell is, the more species are available to reconstruct climate. No botanical information is available in the black grid cells. Species records are most abundant in South Africa, Swaziland and Lesotho.
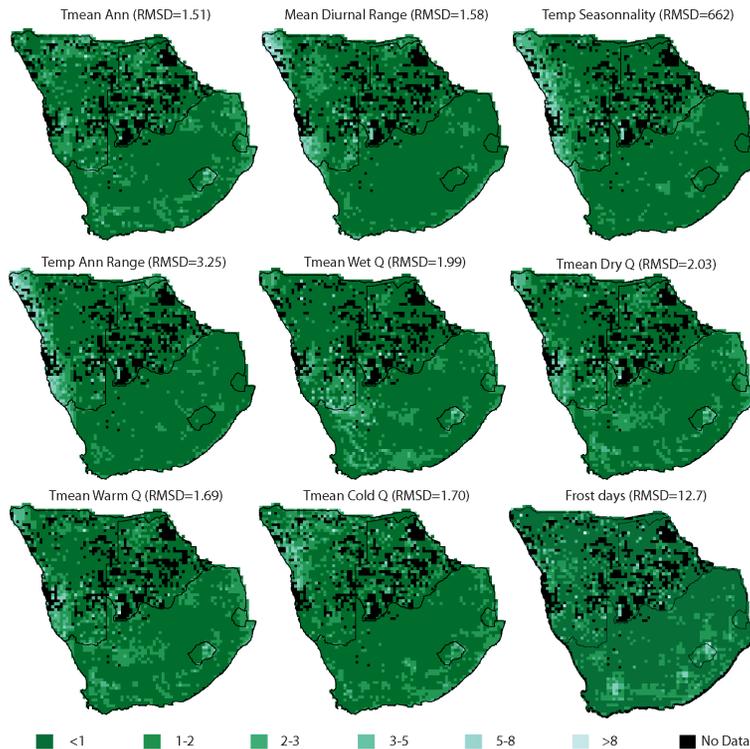
Fig. 5. Geographical distributions of the normalized anomalies of the reconstructions of temperature-like variables (Eq. 12). The scale is identical for all the maps, in units of RMSD. No vegetation information was available from the black pixels.

Fig. 6. Geographical distributions of the normalized anomalies of the reconstructions of moisture-like variables (Eq. 12). The scale is identical for all the maps, in units of RMSD. No vegetation information was available from the black pixels.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper
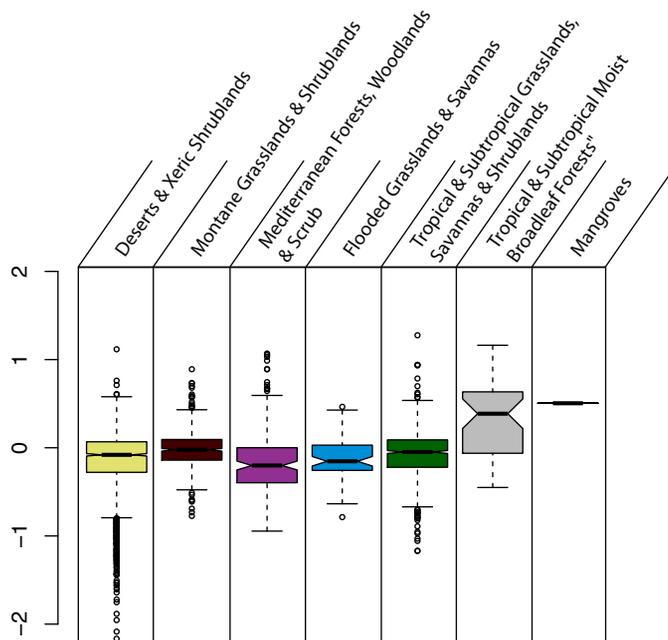
Fig. 7. Boxplots representing the dispersion of the normalized anomalies (Eq. 12) for each biome. There is more dispersion within each biome (length of the boxes) than between, confirming the results of the Between-Groups PCA (90 % of variance not explained by the groups). The colors correspond to the gradient selected in Fig. 3.
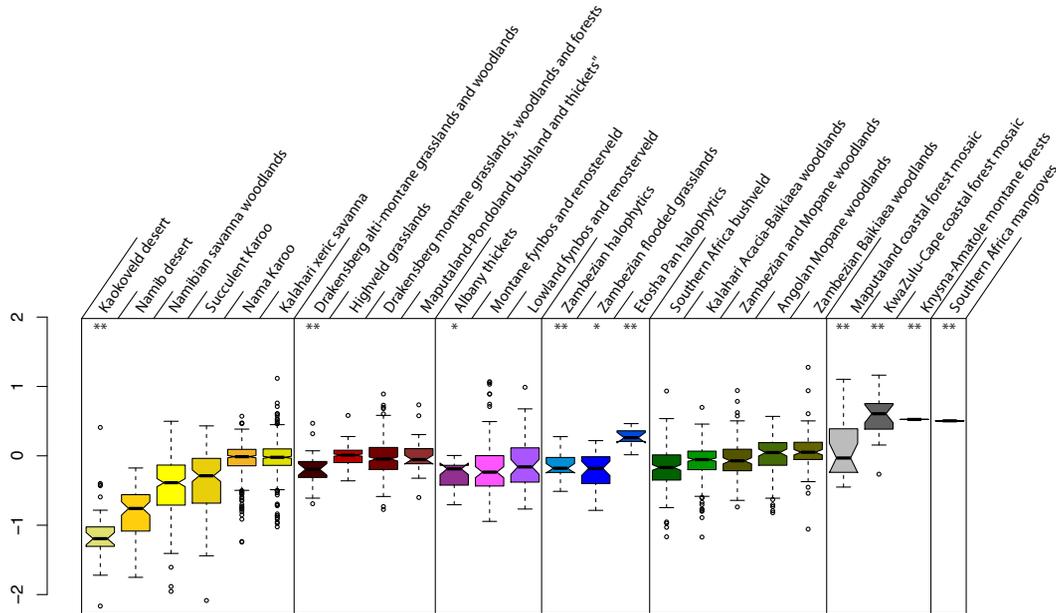
Fig. 8. Boxplots representing the dispersion of the normalized anomalies (Eq. 12) for each ecoregion. There is globally more dispersion within each ecroregion (length of the boxes) than between, confirming the results of the Between-Groups PCA (75 % of variance not explained by the groups). The colors match those of Fig. 3.
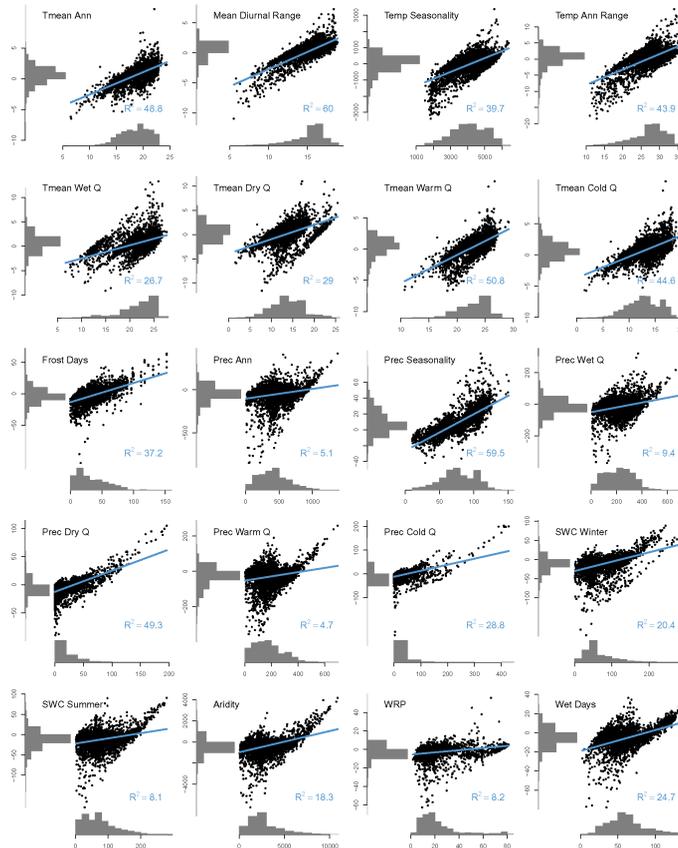
Fig. 9. Anomalies plotted against their targeted values. The density of points is heterogeneous: being very dense around the median climate and sparser at the extremes. This is illustrated by the histograms that represent the marginal distributions. The anomalies are smaller for the best-represented climate values, and increase with distance from the median climate. The blue line represents the linear model fitted, with its associated R^2.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
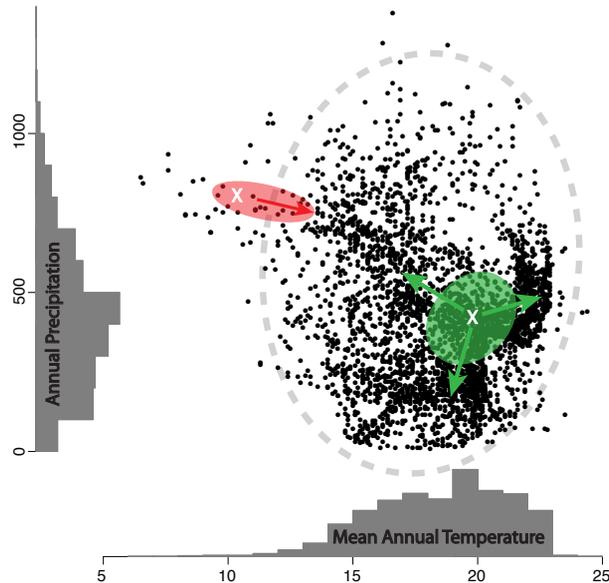
Fig. 10. Scatterplot representing a 2-D-projection of the climatic space of southern Africa for the 2 variables Tmean Ann and Prec Ann. In green and red are the modern positions of two fictious paleoarchives. Those two points represent two very different situations relative to the climatic space: well-represented (green) vs. rare (red) climate. Reconstructing climate changes for the green paleoarchive should be more accurate because it can "move" in several directions around its modern climate. On the contrary, the only major direction the red sample can move to, is towards warmer and drier conditions. Colder temperatures should be "reconstructible" but with an amplitude that may not reflect actual variability.