COMMENTS FROM EDITORS AND REVIEWERS
Replies from us in black
Changes to do in the revised manuscript

Note that referenced lines refer to the line numbers in the original submission.

Reviewer 2:

C Quality of methods/correctness of mathematics: H/L* (L*application procedure of the CA unclear (species, genus, family data?) supplementary data (ES) not available)
Supplementary data were submitted together with the original manuscript. Reviewer 2 could have asked for it by contacting the editor.

Major points
General:
Several comments provided by Reviewer 2 are due to the fact that the reviewer had no access to the ES that we submitted separately during the initial submission process, and did not find the time to ask the editor for a copy.

1) Application of the CA and considerations on its climatic resolution (cf. manuscript title, chapter "Can CA reliably reconstruct MAT or changes in MAT", line 495 and following)
Although the present study is useful to unravel inconsistencies and erroneous entries regarding climatic requirements of extant taxa cited in the Palaeoflora data base it is not qualified to draw conclusions regarding the resolution potential of the Coexistence Approach in general:
RESPONSE 1.1: THIS is the main objection of Reviewer 2, although not further explained. Unfortunately, the present study is qualified to draw conclusions about the resolution potential of the CA (see reviewer 3). We find that increasing the *reliability* (by correcting entries for NLR climate data) decreases the *resolution* of a CA reconstruction. This result is based on numerous single-case studies, all presented in the manuscript and the accompanying ES. Therefore, the CA *in general* will not work because it will give far too wide coexistence intervals when using corrected climate data. Thus, although the reliability can be increased by correcting primary data (climate requirements of NLR taxa), the resolution cannot.

The reliability of the results obtained with the CA and hence the quality of the climate reconstruction depends on and increases with the number of taxa that may coexist. (Mosbrugger and Utescher, 1997).
RESPONSE 1.2: As far as we know, there is no study that has shown this. In contrast, the present study demonstrates that this is not true. The number of taxa that *do* coexist do not have such effect (Tables 1, 4, 7, 8; ES 1, 3, 5, and 6). Increasing the number of taxa only assures that the climate interval reconstructed with CA/PFDB converges to a MAT of c. 16°C, independent whether the real climate of real-world floras or artificial floras are used (ES 6; lines 495–553).
To illustrate this, see Figure 4 of the present manuscript: The quality of the climate reconstruction does not show any correlation with the number of coexisting taxa. Using

climate data from PFDB it is true that when 120 and 112 taxa co-exist the narrowest coexistence intervals are obtained. Please note, however, that these intervals are completely erroneous although they are still in the range of lowland mid-elevation vegetation. At the same time, using corrected data for the same number of co-existing taxa do not show narrow intervals at all, while all intervals touch or almost touch the actual MAT curve (for explanation why they sometimes do not touch see M&M section in the present manuscript). Figure 5 of the present manuscript, using climate data from PFDB, again shows that although the coexistence interval is becoming narrower when more coexisting taxa are included in the analysis, the quality does not increase.

This does not necessarily mean a low width of the resulting CA interval. Therefore a first, crucial step is to provide climate data for all taxa present in the flora, or for all known NLRs, respectively (cf. Mosbrugger and Utescher, 1997).

RESPONSE 1.3: True, a wide CA interval (based on correct data of the used NLR taxa) may be highly reliable. However, at the same time, it may not be very useful, because of the low resolution. This IS about the "resolution potential of the CA".

As far as can be seen from the CA literature, this "*crucial step*" is not a standard procedure (see again ES 2; but see also Xia et al., 2009 and Jacquet et al., 2010 for fully documented studies that fulfil this "*crucial step*").

In lines 154-156 of the present ms it reads: "..even rich (Georgien) floras with numerous recorded taxa translate into a limited number of climatic active taxa". Consequently, the authors use only a small proportion of taxa, e.g. in some of the Georgian floras only 30 % of taxa can be used to calculate temperature, and in the case of the less diverse North American beech forest (Table 7), in one third of the cases the number of taxa contributing with data is close or beyond the critical limit of the CA (10 taxa).

RESPONSE 1.4:

For clarity, we replaced the term "climatic active taxa" with "NLR taxa represented in the PFDB".

The Georgian stands were used to evaluate the precision of CA/PFDB; only taxa were used that could be found in the PFDB. For the possible effect of using many co-existing taxa, see the Chinese floras (cf. response above).

Please also note that the belief expressed here that 10 taxa (or close to 10) represent a critical limit has never been substantiated. Many CA-based papers draw conclusions from the co-existence of few taxa (e.g. Böhme al. 2007; Bruch et al., 2006; Bruch and Zhilin, 2007; Ivanov et al., 2007; Poole et al., 2005; Syabryaj et al., 2007; Utescher et al., 2007, 2009; see ES 2). In a more recent study, Utescher et al. (2009) state, with reference to the original paper, that (p. 107) "*CA commonly provides reliable results when **more than 8 taxa** … In some high-resolution samples the number of taxa used in the CA procedure may fall below **this** critical value (Appendix 2)*" (see also Utescher et al., 2007, p. 145). In Mosbrugger and Utescher (1997) the best resolved MAT interval for a modern flora, which assumedly comprised the real MAT relied on 13 NLRs (Okefenokee swamp). The flora with the most NLRs (32) resulted in the poorest resolved interval (ΔMAT = 5°C).

In addition the authors arbitrarily switch taxonomic levels in cases where species data are not available in the Palaeoflora data base (cf. lines 88-91), use climate data not only for the genus but also for the family level that have by far larger climatic ranges when compared to species.

RESPONSE 1.5: Nothing was "arbitrarily switched"; we used the NLRs that are available in the PFDB. The procedure was explicitly explained in M&M, lines 88–114. This is the most common procedure in CA-based studies (see, for example, Uhl et al., 2006, Bruch & Zhilin, 2007). Please note that several family-level NLRs recorded in the PFDB (Cycadaceae, Dipterocarpaceae, Icacinaceae, "Palmae", "Taxodiaceae", "Cupressaceae") have recorded MAT tolerances that determined the lower (e.g. Ivanov et al., 2007) or upper MAT boundary (e.g., Figueiral et al., 1999; see ES 2). See ES 4 regarding what "*far*" in "*…by far larger climatic ranges*" of related recorded species-level, genus-level, and family-level NLRs really means (ironically: some genus MAT tolerances do not comprise tolerances of all species recorded for that genus in PFDB; see ES 4).

General comment to using species-level, genus-level or family-level NLRs: A repeated critique by Reviewer 2 is that we did not report the number of family, genus, and species-level NLRs (which is, by the way, not true, see specific response above) and "*arbitrarily switch*" between them. It has often been claimed (e.g. Mosbrugger & Utescher, 1997; Bruch et al., 2006, Bruch and Zhilin, 2007), but never proved, that CA based species-level NLRs provide higher resolution than genus-level NLRs. By far the most NLRs used for CA are genus-level at least from what is documented in the literature. As far as documented in the very same literature, the genus- or even family-level NLRs determine the lower and upper boundaries of the MAT (examples provided in ES 2) and not occasionally included species-level NLRs (if one does not count monotypic genera as species-level NLR). It is trivial that if perfect climate data of modern species were available; and only species included, the accuracy of the CA approach could increase (compare Fig. 6/ ES 6 with Fig. 4/ ES 5). However, it is superfluous to estimate modern climates with CA; such parameters can be measured directly. The CA is used to reconstruct past climates, hence, to validate its potency, one should use modern sets of taxa that include taxa that are potential NLRs to represent fossil floras. And these are, in by far the most cases, genus-level NLRs and not species-level NLRs (see ES 2 for examples from CA-based reconstructions that provided a list of NLRs used for analyses; and ES 4 for NLRs recorded in the PFDB with relation to earlier analysed fossil floras). This issue could be discussed excessively, but relates to the general problem of taxonomic ecological specificity. One has to keep in mind that the NLR assumption is much easier established at the genus-level than at the species-level. It appears that fossil taxa that can be compared to particular modern species thrived under substantially different ecological condition during the Cainozoic. This can be evaluated by investigating (qualitatively!) the overall composition of fossil assemblages in combination with sufficient phylogenetic frameworks for as many as possible taxa of fossil assemblages. So far, such frameworks are only available in very rare cases (e.g. *Fagus*, Denk and Grimm, 2009). To illustrate this, we analysed a previous study by Uhl et al. (2006), recommended by reviewer 3, addressing the problem outlined above (see new Fig./Table/ ES).

The electronic supplements that might shed light on this problem (proportions of species/genus/family temperature among the taxa that are "active" in the analysis, as the authors call it) is not available under the cited link (http://www.palaogrimm.org/data/GD11RPP_data.zip).

RESPONSE 1.6: "Climatic active" taxa refers to the reduced set of NLRs used for analysis in contrast to the set of all NLRs that would have been potentially available (see e.g. Uhl et al., 2006; Bruch and Zhilin, 2007, and ES 2; the reasons why some NLRs are replaced by others or omitted, or included despite representing relict taxa, are not clear).

The link will become available after acceptance of the manuscript. Please note that the full ES was submitted together with the original manuscript; if Reviewer 2 did not have it at hand he/she could have asked for it.

Although restrictions regarding taxonomic resolution may occur in CA applications on fossil materials the present test conditions do not allow to assess or evaluate the possible resolution potential of the CA, and hence, the reliability of all conclusions concerning the climatic resolution of the CA drawn in this manuscript are doubtful (e.g. chapter "Can CA reliably reconstruct MAT or changes in MAT", line 495 and following).

RESPONSE 1.7: This comment is difficult to follow. See also general comment of Reviewer 3. We document the performance of CA using PFDB-recorded and more realistic intervals based on ***all*** principal habitats occupied by *Fagus* in North America and Japan, covering a MAT range from 3–20°C, ***and*** various Chinese forests from the farthest north to the nearly tropical south of China, from ***sea-level*** up to high-montane areas. We used the NLRs that were available in the PFDB. We are unaware of a similar (or any) validation that shows that the original hopes about the resolution potential of the CA expressed in Mosbrugger and Utescher (1997), regarding MAT or other climate parameters using PFDB were justified, i.e. allowing to reconstruct temperatures with a precision of 1–4°C. See also our **Response 1.1**, **1.5**

Provide information on the proportion of species/genus/family data in each reconstruction.

RESPONSE 1.8: This information can be extracted from the according ESs, for both our validation data and revisited fossil data.

2) Definition/identification of outliers (cf. page 10, 2nd paragraph, pages 19-21)

The authors question the potential of the CA to detect climatic outliers in climate reconstruction but obviously take a wrong definition of "outlier" as a basis. The examples given on page 10, 2nd paragraph, and on pages 19-21 are not suitable to address this subject because in both cases, 2 equally probable (alternative) solutions are obtained, a problem that usually is critically addressed in CA literature. According to Mosbrugger and Utescher (1997) an outlier is characterised by climatic requirements differing from the demands of the vast majority of taxa recorded in a flora.

RESPONSE 2.1: The procedure how climatic outliers are commonly identified by CA does not match the definition given above. According to Mosbrugger and Utescher (1997), a climatic outlier is recognized by the CA *a posteriori*, and this appears to be the standard procedure in CA-based studies (see e.g. Böhme et al., 2007, fig. 6; see also literature referred to in ES 2). If not all taxa have overlapping tolerances for a tested parameter, the set of taxa is chosen that includes the highest amount of coexisting taxa ("*… that allowed the majority of considered plant taxa to exist at that location,"* Utescher et al., 2007, p. 144). In extreme cases (see text, ES 2) this may mean that an alternative interval favoured by one cold outlier is dropped in favour of an interval to accommodate two taxa that otherwise would be warm outliers (see ES 2). In the example of Böhme et al. (2007, fig. 6-2), four cold-preferring NLR cannot coexist with two warm-preferring NLR out of a set of 42 NLR. Thus the latter are considered climatic outliers, although they can coexist with 36 NLR. This surely does not justify the statement "*differing from the demands of the vast majority of taxa*" (see also data provided by Figueira et al., 1999; Ivanov et al., 2007; Syabryaj et al. 2007; ES 2). The reasoning why the inclusion of 4 cold taxa and exclusion of 2 warm taxa leads to a more correct estimate than vice versa is entirely obscure.

Many climatic outliers could easily be recognized *a priori* and excluded from the analysis: for each taxon one could establish the pairwise coexistence probability p by dividing the number of taxa with climate tolerances that overlap with the taxon's climate tolerance by the total number of NLR minus 1. In case $p = 1$, the taxon could theoretically coexist with all other taxa. One could make a significance test and define a threshold which taxa would need to be excluded, a conservative and easy to argue approach would be to use only taxa with $p = 1$. In the above mentioned example of Böhme et al. (2007, fig. 6-2) this would mean that both warm and cold-preferring potential outliers would have to be excluded from the analysis. Given the data available to us, only two taxa would have $p << 1$ in fossil floras regarding MAT: one is *Dalbergia* (e.g. Utescher et al., 2009), for which the PFDB records a highly erroneous MAT interval (see Table 5), the other "*Pinus sylvestris type"* (e.g. Ivanov et al., 2007). Based on more realistic tolerances, by far the most taxa of the subtropic and temperate zone that are potential NLRs for fossil taxa approach $p = 1$.

3) Validity of "correct" or "real world" climatic data presented in this study, values derived in an extreme orography

Figure 4 of the ms reveals an interesting relation between air temperatures (~ corresponding to the black line) and temperature of the vegetation
RESPONSE 3.1: This is a misunderstanding by Reviewer 2 (see Response 3.3.)

This phenomenon, even not discussed by the authors, exemplifies that climatic limits derived in mountain areas introduce a considerable uncertainty when defining MATmin of a taxon (according to fig. 4 MATmin would be estimated too low, probably by several degrees Celcius).
RESPONSE 3.2: This statement is not comprehensible.

As factors, exposition, microclimate, snow cover, region-specific lapse rate etc. play a role (cf. also works in the Alpine area of C. Körner et al.). Therefore, temperature requirements derived from stands of the higher altitudinal areas provide no probable solution when reconstructing palaeotemperature from fossil lowland communities.

RESPONSE 3.3: There is no such relation or "phenomenon" revealed in Figure 4. The reviewer's differentiation between "air temperatures" (*T*) and "temperature of the vegetation" ($T_{veg,}$), probably referring to the *reconstructed* MATs, is inapprehensible: climate tolerances of modern taxa, which are used to reconstruct MAT intervals, are estimated based on nearby climate stations.

What is shown in Fig. 4 is *T* at different elevations based on a representative set of WorldClim grid cells that cover the Shennongjia area, and the reconstructed *T* based on the taxa occurring at this elevations using CA/PFDB (completely erroneous red bars) and CA/***conservatively*** corrected data (green bars; see M&M for explanation how climate data provided in Fang et al., 2009, were corrected). That the actual *T* (= MAT) can be lower than the lower boundary of the reconstructed *T* demonstrates that our corrected tolerances are still too conservative explicitly stated and discussed in lines 566–583 (cf. ES 3, ES 5, Table 5). Climate tolerances for Chinese genera and species, also listed in PFDB as NLRs, may be 1°C or more ***lower*** than the conservative estimates used for CA/corrected reconstructions. Again, note that all corrected MAT tolerances are based on the original data in Fang et al. (2009) only corrected for such species that occur in counties with extreme differences in elevation using the altitudinal range in the Flora of China (Figs 1-3, ES 8). The altitudinal range of potential NLR in the validation floras (ES 3, 5) has not been taken into account to correct the data in Fang et al. (2009) to prevent circular reasoning!

Please, also consider the consequences of the assumption expressed by rev 2 above: let us assume that modern taxa used as NLR have a distribution that frequently includes mountain habitats with ***different*** climatic parameters than their fossil counterparts found in lowland assemblages. How would a CA then theoretically be possible as the most important NLR assumption is that the fossil taxa have essentially the ***same*** climatic tolerances than their modern relatives? Or does the reviewer suggest that climate tolerances for potential NLRs should, in contrast to Mosbrugger and Utescher's (1997) recommendation, only be based on the lowland distribution of such taxa? This would make it impossible to get climate tolerances for many Chinese genera used as NLRs in CA applications, which often are confined to higher elevations. At the genus-level, all species then must be excluded that only occur in mountains. But how can one be sure that fossil members of such a genus are more closely related and have the same climatic tolerances than the modern lowland representatives and not their mountainous counterparts? In the case of palynofloras, the CA applicant would also need to distinguish between *in-situ* sedimented pollen (autochthonous) and pollen that flew in from nearby higher altitudes (allochthonous).

4) Considerations concerning the validity of palaeoclimate reconstructions (page 15)

The attempt made by the authors to relate climatic equability reported by various authors for several past time slices (e.g. references cited, lines 344-346) to a mean of MAT tolerances over all taxa with climate data available in the Palaeoflora data base (including considerations on number of data entries concerning taxa restricted to very warm climates) is based on erroneous assumptions …

RESPONSE 4.1: We did not make any assumptions, just refer to a number of observations. Namely (1) that the Chinese (and other) modern validation floras result in a MAT interval of c. 15–17˚C with CA/PFDB, regardless of the real climates at various stands. (2) The same interval is found in many reconstructions using fossil floras, with comparable sets of NLR. This is coupled to (3) that most of the revisited NLRs, i.e. such NLRs that are relevant for the revisited fossil floras and crucial to define MAT intervals (see ES 2; ES 4 for the full list of revisited NLRs) can coexist at around 16˚C.

… and therefore has to be regarded as invalid: Palaeoflora does not contain any representative fractions of taxa covering the various Koeppen climate types. The data base grew in adding new records according to the demands, the NLRs listed for fossil sites studied in the frame of co-operations with colleagues (in each case aiming at a complete set of climate data for all known NLRs). The fact that over 80 % of the fossil floras analysed so far represent sites from younger Palaeogene to Neogene lowland areas of the mid-latitudes, alone can explain the frequency distribution discussed on p.15.

RESPONSE 4.2: It is not relevant to the study whether the Palaeoflora comprises such "representative fractions" or not. PFDB comprises most of the taxa used to reconstruct past climates (e.g., Oligocene to Pliocene of western Eurasia). Hence, to validate the performance of CA/PFDB these taxa have been used.

Furthermore, the authors do not present any novel insight in stating that gradients constructed from means might be to shallow (cf. procedure used in Utescher et al., 2011; Klotz, 1999; Pross et al., 2000)

RESPONSE 4.3: We explicitly state that our data confirm the observations of Klotz (1999) and Pross et al. (2000). Nevertheless this long-known insight did not affect the critical use of "centre values" (which hardly represent "means" in a statistic sense) as the basis for many conclusions in CA/PFDB-based studies (see e.g. references and documented data compiled in ES 2; see also "oscillations" of Utescher et al., 2009, fig. 7, and variations/slight shifts elsewhere, e.g. Bruch and Zhilin, 2007).

Minor comments

Lines 40-41

Actually, there are two NECLIME synthesis volumes (Bruch et al., 2007; Utescher et al., 2011). This has to be updated.

RESPONSE: We re-phrased.

Line 49

"..with an assumed precision of 0.1 °C.."

As far as I know, such a devious precision in temperature reconstructions using the CA was never claimed in CA related literature. Therefore, this statement should be removed or substantiated by a reference.

RESPONSE: CA studies provide MAT interval values down to the first (see ES 2) or even second decimal (e.g., Uhl et al., 2006; Bruch and Zhilin, 2007, table 3). It is common practise in natural sciences to only use the number of decimals which are considered to be precise or at least meaningful. If authors do not assume that reconstructed intervals have a precision of 0.1°C but only of full degrees it makes little sense to write MAT = 15.6–17.6°C instead of MAT ~ 15–18°C. It makes also no sense to discuss results using these decimals if it is assumed that the reconstructions are not that precise. For instance, Bruch and Zhilin (2007), who found a MAT of "*around 15°C*" (lower MAT boundaries: 13.3–15.3°C ; upper MAT boundaries 15.6–16.9[–20.8°C]) write (p. 46), referring to the study of Syabryaj et al. (2007): "*Characteristically, the flora from the Ukrainian Plain provides slightly cooler temperature values (MAT: 15.6–15.6 °C; CMT 2.3–7.1 °C, WMT 25.3–26.3 °C) than the flora from the Carpathian Basin (MAT: 15.6–16.1 °C; CMT 6.6–7.8 °C, WMT 25.4–25.6 °C), especially in winter temperature…All other information available from European records shows warmer and especially more equable conditions with low seasonality. From Central Europe, data of Aquitanian floras from Hungary are published by Erdei et al. [2007], yielding a MAT of 15.6–18.8 °C, CMT of 5–10.2 °C, and WMT of 25.6–27.5 °C. Further south, data from Turkey (Akgün et al., [2007]) provide even warmer values of MAT 16.5–21.3 °C, CMT 5.5–13.3 °C, and WMT 27.3–28.3 °C. Thus, all data available from European floras give much warmer mean annual temperatures (all >15.6 °C).*" Rounded to full degrees, one could only say that *all* discussed fossil floras addressed here could have thrived at a MAT of 16–17°C.

Lines 72-75

The statement that "..wrong climate intervals of extant taxa do not seriously bias the analysis.." cannot be substantiated by the cited reference and therefore this reference has to be removed here. And as far as I know there exists no other CA related publication that would support such absurd argument the authors afford here. To clarify this I include a citation from the original paper

"Of course, the coexistence approach also includes a number of shortcomings. It is based on a few assumptions which may all be wrong in some cases (see Section 8) and it depends on the quality of the data base that contains the nearest living relatives of fossil taxa and their climatic requirements".

RESPONSE: We wrote "… following the ***philosophy*** of the original paper (Mosbrugger and Utescher, 1997) and most papers since then ….". Mosbrugger and Utescher (1997) explicitly also note (p. 73–74) "*It is part of the **strength of the coexistence approach** that such errors* [wrong assignment of NLRs, mismatch between climatic requirements of NLRs, wrong climatic requirements recorded] *can, **for the most part**, be detected. … That the data does contain **largely correct entries*** [but see ES 4] *is shown by the successful application of the coexistence approach to various modern* [i.e. <u>four</u> floras from three different climatic settings] *and fossil floras* [i.e. <u>one</u> sequence of fossil beds] *and by the fact that, in both modern and fossil floras, **statistically highly significant coexistence intervals** are calculated in which 88-100% of taxa can coexist. … So far we have illustrated that **all errors, that may occur and which lead to climatic outliers, can be detected** and **will not affect the** determination of coexistence intervals. Another advantage of the coexistence approach is that it is also **robust against minor errors which do not lead to climatic outliers**. … Thus it becomes evident that the coexistence approach **implies some basic errors, but proves to be a robust technique in***

which serious errors can be detected." They conclude this paragraph (p. 74) saying that they "*…expect the coexistence approach to be* ***perfectly applicable*** *in the Tertiary but to produce best results for Oligocene and younger floras*."

Given this optimistic assessment, the errors, which are acknowledged in the citation above, could not have been considered by Mosbrugger and Utescher (1997) to seriously bias the analysis. Indeed, if this risk (as stated by the reviewer) would be imminent, then the conclusion of the paper would have been that the CA may prove to be a reasonable alternative in the future, given that the effect of first-order errors type 1–4 (not only 1–3) on the outcome of the CA is fully studied rather than (p. 76) "*…we consider the coexistence approach to represent a* ***reasonable, robust and the most widely applicable technique of quantitative reconstructions*** *of terrestrial palaeoclimates in the Tertiary*."

Based on this conclusion and the optimistic tone of the paper despite apparent problems, most later authors did not consider that such an error could seriously bias the analysis: erroneously recorded taxa (for details refer to ES 4) determined many MAT intervals in numerous studies (a selection is provided in ES 2). In none of these studies the authors considered the possibility that the obtained MAT interval shifts, or the fact that some taxa were recognized as outliers, could have been due to erroneously recorded climate tolerances of NLRs, i.e. error-type 4. For instance, Utescher et al., 2009, explicitly state that "… *shifts to higher* [temperature] *values* [MAT ranges from 12–18°C to MAT > 15.5°C] *basically are due to the occurrence of Engelhardia* [MAT_{min} = 15.6°C in PFDB] *and/or Reveesia* [MAT_{min} = 17.2]*.*" (p. 107; appendix of this study is included in ES 2). Two NLRs with too conservative intervals: for both *Engelhardia* and *Reveesia* (occurring ≤ 1500 m) the PFDB tolerance is at least 3–4°C too low (Fang et al., 2009, corrected; Xia et al., 2009; cf. ES 4). Thus, a type 4 error (according to Mosbrugger and Utescher, 1997, section 8) could be the simple reason for the observed "shifts" (notably only the lower boundary shifts >2°C; cf. ES 2). Thus, we feel that our statement is well substantiated by the tone of the original publication and works based on it.

Lines 119-121

The statement "..could not be validated because of the non-dissemination policy of the NECLIME group" is most embarrassing when considering that the second author was part of NECLIME for several years. Instead of contributing with his expertise he decided in 2009 to return his membership and explicitly asked to be removed from the NECLME mailing list and data distribution.

RESPONSE: The second author was never part of NECLIME. As a matter of fact he does not know how his name came on the list of NECLIME members and on the mailing list. When he found out in 2009, he asked for the correction of this error.

I have information that the partial citation from personal email refers to an inquiry of the authors concerning climate reconstruction from the floral record of Iceland. T. Utescher offered a complete set of climate data in exchange of a list containing fossil taxa and corresponding Nearest Living Relative(s). However, the authors were not interested in a two-way exchange. I recommend to remove this quotation, otherwise cite in full length.

RESPONSE: A list of NLR taxa was supplied directly after we received that mail (the file was created on 29/09/2010 and subsequently sent) along with the promise that as soon as the

comparison PFDB data versus our corrected data would have been made by us, the outcome would be provided to T. Utescher together with detailed information on fossil taxa-NLR associations and alternative climate parameters. However, we never received any reply to that mail and turned to other projects.
Following the editors suggestion this part has been entirely deleted.

Line 137
The authors state that the use altitudinal ranges cited in the Flora of China in order to correct climate data published in Fang et al. (2009). However, the Flora of China provides no altitudinal information at county level, at least as far as know. To improve the reliability of the present data reconstruction a more detailed explanation should be added how this was done.
RESPONSE: The exact description of the procedure was unfortunately lost when shortening the paper. We re-included it.
The data used to correct the data of Fang et al. (2009) is provided in ES 4, 7, 8. The archive also includes DIVA-GIS generated files based on the WorldClim Grid by Hijmans et al. (2005). ES 4 also lists exact details on the source of our corrected data.

Lines 140-141
ES 1-8
The electronic supplement is not available under the link provided by the authors
RESPONSE: See above.

Lines 175-178
This statement cannot be substantiated by the original paper and has to be rephrased. Mosbrugger and Utescher only state that the CA potentially recognizes outliers: "generally, the higher the diversity, the higher the climatic resolution and the greater the chance that inconsistencies or 'outliers' can be recognized"
RESPONSE: Please, read the complete section 8, which already has been partly cited above (response to l. 72–75). "*Misidentification of a fossil taxon or* ***poor modern climatic data*** [this would be error type 4] *can be an additional source of error. It is part of the strength of the coexistence approach that such errors can,* ***for the most part****, be detected…. For example let us assume that the data base contains some errors of type 1, 2, or 3* [notably type 4 is not mentioned, but of course also modern climatic data of not so good quality could inflict climatic outliers as shown in this paragraph]. *When the coexistence approach is applied to numerous fossil floras,* ***these errors will rapidly be identified as permanent outliers****, provided … fossil floras of a certain diversity (10-15 taxa) are analysed.* [etc. etc.]" Secondary literature (e.g. Bruch et al., 2006, Utescher et al. 2007, 2009) refer to this study when stating that errors in the data base can be ruled out, because they would be recognized as outliers.
As far as documented in the revisited publications so far none of the taxa compiled in Table 1 and only one taxon compiled in Table 5 represent a *"permanent outlier"*. The genus-level NLR *Dalbergia*. However, no effort was made to revise its MAT tolerance, instead, in a most recent account, Utescher et al. (2009) again explain the fact that *Dalbergia* is recognized as an outlier by the assumption that it had a much broader range in the Tertiary as it was done in earlier studies. According to Xia et al. (2009) the lower tolerance of *Dalbergia* is 12°C, and

we estimated a MAT_{min} of 10°C (see ES 4). This is corroborated by its altitudinal occurrence in the Shennongjia area (cf. ES 5). Based on more realistic data than stored in PFDB, *Dalbergia* would be fully within the range of coexisting taxa (cf. Xia et al., 2009; see also the data provided for *Cassia*). Other erroneous entries in the data bank (cf. ES 4; see also PFDB-independently derived intervals by Xia et al., 2009; Jacques et al., 2010; Wang et al., 2010; ES 2) were also ***not*** permanently recognized as outliers in CA-based analyses (examples provided in ES 2).

Mosbrugger and Utescher (1997) give also some examples for taxa with too narrow or unrepresentative climate intervals (p. 74). "Sequoia, *for instance, forms a 'cold outlier' in most climatic analyses of Tertiary floras. This indicates that the modern, monotypic genus* Sequoia *is restricted to relatively cold conditions whereas its Tertiary relatives probably also lived in warmer habitats. The use of* Sequoia *as a climate indicator has therefore led to misleading results in previous studies (e.g. Van der Burgh, 1973; see Table 4)*" In table 4 in Mosbrugger and Utescher (1997), it is said that Van der Burgh (1973) inferred a ("misleading") MAT of 12°C for the Ville Formation. According to Thompson et al., 1999a, the real MAT tolerance of *Sequoia sempervirens*, is 9.4–15.3°C. This would be very close to the lower boundary or even within the MAT intervals shown in Mosbrugger and Utescher (1997, figs 3–4) for the Ville Formation. On the other hand, other, important errors in the database remained undiscovered. For instance, *Engelhardia*, recorded with a MAT_{min} of 17.5°C at the time of the paper. This was corrected to 15.6°C only later to avoid its placement as an outlier. The genus frequently defines the lower MAT boundary in fossil floras. However, this value is according to Kvaček (2007) too high (MAT_{min} 10–14˚C). Despite this, the taxon is still used to define the lower MAT boundary in many fossil studies with a MAT_{min} of 15.6°C (e.g., Bozukov et al. 2009; Larsson et al. 2010, Liu et al. 2010; cf. ES 2). The genus occurs in China at MAT down to 11°C at least (Fang et al., 2009, corrected; cf. ES 4), so both the original 17.5°C as well as the 15.6°C represent remarkably wrong data. In the case of Liu et al. (2010) the correction of this erroneous entry would change the reconstructed interval from 15.6°–20.8°C to c. 11°C–20.8°C. Omission of *Engelhardia* would result in a CA/PFDB interval of 9.3–20.8°C.

Line 211 and following

When following the original description of the method (Mosbrugger and Utescher, 1997) taxa providing alternative solutions that are equally significant (same number of coexisting taxa) are not denoted as outliers.

RESPONSE: A taxon outside the preferred climate interval is an 'outlier'. Note that a modern flora technically cannot have less than 100% coexistence unless the tolerances recorded in a database are wrong or the taxon is misidentified. The justification for using the coexistence percentage to assess "statistically" the "significance" of alternative intervals based on fossil floras is another issue not addressed in our paper.

Lines 289-290

The studies by Klotz and Pross et al. are not based on artificial pollen floras but on synthetic floras. This has to be corrected.

RESPONSE: We corrected this accordingly.

Lines 496-499

Explain why the CA/PFDB worked well with Georgian floras

RESPONSE: The observation that the *same* (narrow) interval is reconstructed for climatically and floristically significantly *different* stands in Georgia (Table 1 and according chapter) hardly documents that "*CA/PFDB worked well with Georgian floras*".

Lines 509-511

In the majority of cases where MAT discrepancies between Palaeoflora data and data presented in this study are over 5 °C it can be assumed that the data provided by the authors are based on single occurrences of plants or/and are biased by other factors (cf. 3) and thus cannot provide a probable solution in climate reconstruction.

RESPONSE: See data provided in the ES, the compendia by Thompson et al. (1999a, b, 2001) and Fang et al. (2009) and studies that provide independently obtained climate data (MAT tolerances) for NLRs from other sources (e.g. Klotz, 1999; Xu et al., 2008; Xia et al., 2009; Jacquet et al., 2010; Wang et al., 2010) that all show discrepancies > 5°C with the PFDB records.

Mosbrugger and Utescher (1997), on the other hand, do not say that remote occurrences or occurrences of only a single species of a genus should be ignored to estimate the tolerance of a species- or genus-level NLR. It may, however, indeed be worth using only the 10-90 percentiles instead of all available data (see Thompson et al., 1999a, b, 2001; data included in ES 6), but it is difficult to accommodate such an approach with the logic and application of CA. CA reconstruct past climate using the coexistence of taxa assumedly occurring at the lower or maximal tolerances and high resolution depends on the occurrence of exotic taxon pairs.

Line 515 and following

Due to lee effects and large gradients the identification of MAP limits in altitudinal areas is even more critical

RESPONSE: MAP values have not been corrected but are taken directly from the reference literature and gridded climate data as stated. Please note that the climate station data used, e.g. by Thompson et al., 1999a, b, 2001 are as dense as possible. It would, however, be interesting to see how these effects have been dealt with when compiling the MAP limits of modern taxa used as NLRs for the CLIMBOT database, in particular if, as stated, only four to six climate stations were used to establish these tolerances. Furthermore, we would be curious to learn why the reviewer thinks that fossil assemblages are restricted to vast lowland plains with no mountains.

Line 529 and following

As stated above the Palaeoflora data base does not contain any representative fractions of cool, warm temperate or tropical taxa. Therefore, considerations about any data base mean intervals make no sense.

RESPONSE: See response 4.2.

Line 538 and following

Experiments with artificially pooled floras: Explain what you would expect as a result from such a pooled flora, when assuming a proper functioning of the CA.

RESPONSE: We do not assume that CA can "function" properly; therefore we did not expect any other result. However, Mosbrugger and Utescher (1997) assumed that the coexistence percentage can be used to provide a statistic means for the reliability of a reconstructed interval. Our results lines 538ff (full details provided in ES 5–6) demonstrate further that using more realistic climate intervals than recorded in PFDB, coexistence percentages significantly lower than 100% are very unlikely to be found at all (line 542), even if random sets of subtropical and temperate taxa are mixed covering a large climatic range. The purpose of the pooled floras is to demonstrate the 'black-box' effect when CA is used on fossil assemblages, in particular such ones of pollen and spores, that most probably comprise allochthonous taxa.

Lines 595-596

For exotic climates cf. Utescher et al. (2009).

RESPONSE: The combinations of MAT, CMT, WMT and MAP intervals reconstructed for various sediment slices reported by Utescher et al. (2009) are hardly exotic (cf. ES 2). The best resolved intervals (Unit 0.25 m; 1.25 m) concur with the situation found today in Charlotte, NC, and Chengdu, Sichuan. Most reconstructed climates would also include the climate of North Iran (area around Rasht; see ES 2); less resolved intervals would be in the range of many modern Cfa climates including some Csa and Cwa climates. However, only results are provided but no lists of fossil taxa and used NLRs, hence, possible exotic climates in Utescher et al. (2009), and their reason, cannot be commented.

Line 599 and following

Regarding the Cenozoic floristic record the list of taxa indicative for warm and cool conditions, respectively seems to be far from complete. Again, the wide ranges presented include conditions that can be regarded as unlikely (e.g. when making a reconstruction for a fossil lowland flora).

RESPONSE: The full list is provided in the ES. Regarding which NLRs are meaningful and which not, see responses above.

Lines 608-610

It might be true that the data the authors use "upper boundary, and accordingly the center value, provides little or none information". Obviously, they missed genera distributed in the subtropical and tropical area (MAT up to ca 32 °C) and did not allow for climate data sets for species that may have very significant upper limits.

RESPONSE: We did not miss these genera, they simply do not occur in the validation floras nor do they occur in the revisited fossil floras that have been used to reconstruct past climates. Our conclusions are based on floras of climate zones that have been in the focus of most CA applications, i.e. mostly warm-temperate and subtropical floras. These are also the floras we are interested in.

The highest MAT that we encountered in the surveyed literature is significantly lower than 32°C, and, except for one case, the reconstructed intervals are broad (ΔMAT ranging from 7.4–12.7°C; lower MAT boundary 13.8–16°C; upper MAT boundaries 24.2–26.5°C; cf. ES 2). Böhme et al. (2007) reconstructed a MAT of 22.2–24.2°C for the late Ottnangian of Central Europe based on wood remains (20 NLRs). The lower boundary in this case is defined by the joint genus-NLR *Xylocarpus*/*Carapa* (the upper interestingly by the non-tropical genus *Castanea*), which also occur in younger strata, but then are recognized as ‘climatic outliers’ (together with the allegedly tropical [see Ghazoul, 2011] family-level NLR Dipterocarpaceae) as their recorded tolerance is outside the upper limits of four taxa in that flora (*Crataegus, Cupressus, Robinia, Pistacia therebinthus,* however, not of that of *Pistacia* also included in the list). The reasoning for this is entirely unclear from that paper.
It may (theoretically) be that in the case of very hot, tropical floras of Africa or South America (MAT of 32°C are not found in North and Central America, Eurasia and Australasia, the maximum MAT here is 28°C (ES 7; Lieth, 1999; Hijmans et al., 2005), which is also the cut-off for many taxa listed in the PFDB), the upper boundary will have more relevance, but, to our knowledge there is no study to substantiate this hypothesis. The highest ‘centre values’ (> 20˚C MAT) and upper boundaries in the revisited literature were (ES 2)

Lines 615-618
Based on the analysis of modern floras Klotz (1999) demonstrated that the position of the real temperature within the CA interval is related to climate and latitude.
This contrasts the results presented here and should at least be mentioned. I assumed that the coincidental position of measured data with respect to the CA interval shown here is caused by the arbitrary shifting from species/genus/family data the authors perform.
RESPONSE: The analysis of modern floras by Klotz (1999) confirms the results presented here. We added the missing reference.

Lines 618-623
Considerations about the resolution the authors obtain from the application of the CA - even when using "corrected" data - cannot be retraced by the reader if do not clarify how many species/genus/family data in each case take part in the analysis.
RESPONSE: This information can be retraced from the comprehensive ES.

Lines 643 and following
To assess palaeoclimate conditions for a fossil flora the authors propose to use climatic data of modern validation floras and link those to standardized vegetation types. I think that such an approach would not work because there is broad agreement that a 1:1 relation of modern and Cenozoic phytocoenoses cannot be established.
RESPONSE: Therefore we wrote “*…could **possibly** allow…These could then **perhaps** be used…or **semi-quantitative** signal.*” If this is not possible or feasible, the CA can be considered plain dead. Since the reviewer shares our doubt about the possibility of such an approach we deleted the according sentences.