Clim. Past Discuss., 8, C2143–C2176, 2012
www.clim-past-discuss.net/8/C2143/2012/

Climate of the Past Discussions

Interactive comment on "On the differences between two semi-empirical sea-level models for the last two millennia" *by* M. Vermeer et al.

M. Vermeer et al.

martin.vermeer@aalto.fi

Received and published: 1 November 2012

Author reply to Aslak Grinsted

1 November 2012

Review of "On the differences between two semi-empirical sea-level models for the last two millennia." By Vermeer, Rahmstorf, Kemp and Horton (Clim. Past. Discuss.)

This discussion paper compares two semi-empirical sea level models with proxy records of sea level rise. The desire of the authors appears to be to argue that the authors own model (K11) is better than ours (G10). One of the major problems with this work is the decidedly biased analysis and presentation. Every possible deficiency of the G10 model is highlighted, whereas deficiencies of the K11 model and the proxy sea level record are downplayed or outright ignored. One example is in the introduction of the two models where the G10 model is called unphysical, whereas the obvious unphysical infinite response of the K11 model is ignored.

One very problematic aspect of this work is that all the tests are against proxies of local records of sea level whereas the models are of global sea level. For example, the Kemp et al. North Carolina record has a very high 20th century rate of rise compared to tide gauge based estimates. Kemp et al. (2011) has a figure that demonstrates that there is still considerable disagreement between millennial scale proxy records of sea level. Clearly it is too early to calibrate semi-empirical GSL models against these records, and any validation should acknowledge the full uncertainty. Indeed B. Horton acknowledged as much in an email to me where he wrote the following on calibrating the K11 model to the NC RSL record:

> - 'If, as you say in your email, it was to point out that "it is premature to calibrate a global semi-empirical model to this single proxy record" then I would have been very supportive. . . indeed it's the basis of our current research and many others within the sea-level research community.' – Benjamin Horton (10. Dec 2011).

The discussion in section 4 does not remove this rather fundamental problem. It is still premature. I find several critical issues with the work:

- Over-confidence in the NC RSL record as a record of Global Sea Level. The salt marsh proxy records are local and not global sea level reconstructions. It is simply too early to calibrate / validate against these records, as there is still considerable disagreement between proxy records from around the globe (see e.g. Kemp et al. 2011). The very high 20th century rate of the NC RSL record is only marginally compatible with tide gauge based estimates. Over-confidence in the proxy record would lead to over fitting of the K11 model.
- Highly biased analysis and presentation. I can appreciate the authors desire to believe that their own model is better than alternatives. However, I am convinced that disinterested parties would agree that it comes across as biased and unfair.

Finally, the K11 model is the same as the G10 model with a few additional terms. It is therefore no surprise that it can mimic a wider range of behaviors and fit a greater range of curves. I do not need a study to show me that. However, I remain unconvinced that the additional parameters can be meaningfully constrained by the proxy data. Further it is clear that there is a limit beyond which the use of a perpetual and infinite response becomes detrimental even with perfect data.

Aslak Grinsted

Dr Grinsted correctly points out that in our considered judgement the K11 model is better for studying sea-level behaviour over many centuries. Our aim was to present an objective and timely comparison of semi-empirical models. The new high-resolution proxy data were not available to G10, but enabled and motivated our present study. We thank Dr Grinsted for the spirit of his fundamental critiques and we sincerely appreciate the extensive, often presentation-related remarks in the Detailed Comments section, many of which we may still put to good use. The core of Dr. Grinsted's remarks remains the same as those presented in his published comment on Kemp et al. (2011). Our position remains unchanged on these issues and are published in our response.

Semi-empirical models need datasets for global temperature and global sea level that extend further back in time than instrumental measurements. Dr Grinsted in his review acknowledges that multi-centential datasets are needed for the necessary calibration. This requires proxy data, the very nature of which is imperfect in its spatial and temporal coverage. We believe that the North Carolina proxy sea-level record does approximate global sea level over long time scales within its stated uncertainty. At the time when K11 was developed it was the most appropriate dataset for us to use, but future work may very well supersede it. It should be noted that this assumption also lies at the heart of using longer compilations of tide-gauge records, which prior to 1850AD are from just four sites all in western Europe. Similarly it is necessary to accept the assumption that whatever temperature reconstruction is used is representative of the global mean when proxy (and instrumental) datasets show spatial bias in their distribution. As a scientist whose research focuses on proxy-based reconstructions Prof. Horton characteristically wants additional datasets and continually seeks to address this issue. We were disappointed to see a partial quote from a personal email cited as "evidence" in professional scientific discourse.

GENERAL COMMENTS

One philosophical issue I have with the comparisons in this manuscript is that the K11 model was formulated with prior knowledge of the Kemp et al. proxy record in mind, whereas the G10 model was not. I.e. we would expect K11 to fit this dataset better regardless of whether it really improves the model. In such a case there has to be a very high bar for choosing the more complicated model over a simple model which was formulated without this prior knowledge. This is hard to address, but this issue should be acknowledged in the manuscript. The paper would also be more balanced if the Vermeer and Rahmstorf (2009) and Rahmstorf (2007) models were analyzed with the same scrutiny as the G10 model. These models may not have been designed for long hind casts, but they were predecessors to the K11 model, and were formulated without prior knowledge of the NC RSL validation target.

The K11 model was developed with the North Carolina data set. At its core however it is a standard multi- time scale response model, and one can test that the added free parameters (practically speaking, τ and $a_1 T_{0,0}$) collect only little likelihood around ∞ and 0, respectively. The only parameters that can not be meaningfully resolved from the NC data, are a_1 and $T_{0,0}$ separately (as opposed to their product). We made a conscious decision not to include R07 and VR09 into the analysis because as stated already in R07, the long time scales of sea-level adjustment are not resolved but instead taken as "infinite", so these models are first-order approximations to the initial sea-level adjustment that could not possibly be valid over a millennium. We accept (without evidence) that the R07 model would perform poorly, and the VR09 model is only different from R07 in aspects that have little relevance for the multi-century time scale. The evolution of these models from R07 to VR09 and K11 is an example of scientific progress. R07 purposefully did not attempt to resolve the long time scale tau since he deemed this to be impossible with data spanning only 130 years, while the proxy data now make this possible.

Regarding the perpetual response term: I acknowledge that there is physical motivation behind a slow response term, but I think that implementing it as a non-equilibrating perpetual response is dangerous as it eventually must lead to a too great response. It might improve the simulation of the true processes on timescales that are longer than about 1 to 2 tau, but as you approach the response time of the slow processes that you are trying to capture, then it will start to diverge with a too high rate. This would tend to lead to a positive bias in long term projections made using the K11 model. For long-term projections you may well be right that the Jevrejeva et al. (2012) projections are biased low as the lower range corresponds to response times that are much shorter that 500yrs. The upper range however does cover very long response times.

The K11 model (like all semi-empirical sea level models) is incomplete and a work in progress. We acknowledge that long-term projections made using K11 may include a degree of bias, but we believe that treating a very long finite term as infinite would not cause significant bias. The K11 model performs well over 900 years, suggesting that the slow response time scale is long compared to this; so projections over 300-500 years should be safe. Especially over 300 years, the conservative Schaeffer et al. (2012) choice. Of course any projection made for 2300AD is subject to considerable uncertainty that lies outside of the models themselves and should be treated with an appropriate level of caution.

There is no such thing as a non-informative prior. Please specify exactly what priors you used. I assume that you are using a flat prior for tau so that P([0 100]) = P([100000 100100]) = P([-100 0]). This is obviously a poor assumption as we know that tau has to be positive. You will probably agree that there is a substantial difference between tau=1 unit and tau=100 units, but there is almost no difference between the models using tau=10001 units and tau=10100 units. If you have no information on what units have been used or no information on the typical timescales involved of the process, then it is clear that you should assign a much lower prior probability to tau being between 10001-10100 units than between 1 and 100. A more appropriate choice for parameters such as tau would be a simple flat prior for log(tau).

This (a flat prior for $\ln \tau$) is actually what we used; we should have noted that in the manuscript.

Please refer to Albert Tarantola's book "Inverse Problem Theory" section 6.2 for more detail on this argument.

Why do you calibrate on post-1000AD NC RSL data only? The reasoning behind this subjective restriction of the calibration interval is not presented.

It is not subjective: see section 5, "The discrepancy before 1100 AD".

The K11 model has two terms: One where the response is slowly decelerating over centuries, and another perpetual term which does not slow down. These two terms will have a near identical response over short time scales and it is therefore extremely hard to empirically determine the relative contributions of the two terms without a long calibration target. It must be very close to the detection limit of what you can extract from the uncertain proxy data. I question whether a millennium of proxy data is really sufficient to calibrate the model.

All figures and analysis use a deceptive base line reference. It is at present day (time of proxy collection) that we know what sea level is and numbers are given relative to this reference. Although the semi empirical models are able to absorb any change in reference, then choosing the reference has a big impact on the likelihood function. When the long baseline is chosen then it becomes less important whether the model gets the 20th century rate right. E.g. consider a model where S is constantly zero. With a long baseline, then this would be considered much more likely, than if you chose a recent baseline. This will also mask how VR09 diverge from the NC record within a century (fig2). It is also not clear how the different proxy records in fig4 are aligned vertically. The present day is the natural reference to choose. Indeed this is a critical constraint used to calculate the 2.4mm/yr rate that Kemp et al. estimates for the 20th

century. If you do not apply this constraint on the intercept then you will get a higher 20th century rate, and the marginal "agreement" with tide gauges GSL will disappear.

We do not see this in our formulation: the baseline choice makes no difference for the likelihood function. Both models and sea-level data only account for changes over time, not absolute sea level, so the choice of baseline is arbitrary.

The NZ and Tasmania proxies are aligned on recent sea level, just like Sand Point and Tump Point in North Carolina.

There is an uncertainty in the subsidence correction, and you acknowledge that virtually all proxy data become more uncertain back in time. Yet, the width of the NC RSL confidence interval is exactly the same in 500AD as it is in 1700AD in your figures. That cannot be right. The uncertainty of the observations also enters the likelihood function, and therefore affects all model calibrations. That makes me wonder if you constrain the K11 model much stronger than it should be to the data. Too high confidence in the calibration data would tend to favor models with more parameters.

The confidence interval (orange bands) is generated by the polynomial fit to the raw data that is used to summarize the North Carolina dataset and it is, perhaps surprisingly, uniform over time. This summary was not used in any analysis (as stated in Kemp et al., 2011) but rather is used to aid the reader because showing the individual proxy data points can make graphs look overloaded and hard to read. GIA uncertainty is not part of our uncertainty model but addressed separately. We do take serial correlation into account in our likelihood function.

The K11 hind-cast deviates strongly pre-1000 AD, and I would guess that the RMS error of the original G10 hind-cast is substantially smaller than that of the original K11 model when calculated over the entire NC RSL record. That is remarkable considering that the K11 model was formulated with the specific goal of fitting this record. In figure 4b they show that our original G10-Moberg model has very good qualitative agreement with the oldest New Zealand proxy data with no further calibration. Yet somehow K11 comes across as 'winning' this test in the text.

G10-M08-NZT bifurcates. Immediately before the earliest sea-level data point used there are two alternative, very different solution "branches". The lower branch has a particularly poor fit with the proxy data. Although the K11-MO8-NZT model does not fit the proxy data well, it does have a single, well defined hindcast.

Anecdotically, when testing our implementation of the G10 model, we saw more bifurcations, but never with K11.

There is allowance for any uncertainty due to changes in variability (e.g. tidal range) which will affect salt marsh records. Please make a case for why this is neglible.

The salt-marsh reconstructions are of relative sea level and therefore take into account all eustatic, isostatic, tectonic and local processes. The agreement between the reconstructions from the salt marshes of Sand Point and Tump Point that lie in separate water bodies gives us confidence that local scale processes are not the primary drivers of reconstructed relative sea level. The original Kemp et al. (2011) publication includes a supplementary section on the role that tidal range (and other local factors) play. A scenario where the tidal range at both sites changes by the same amount and at the same time without a detectable stratigraphic signature is unlikely as confirmed by agreement between the two records.

Please document the exact formulation of your likelihood function, as this is critical to all fits. Does it allow for long-range persistent errors (such as would arise from uncertainties in GIA)? Also what is the calibration interval and why was it chosen? Is this choice important for the K11 model? How does the fit look if you calibrate over the entire interval?

The likelihood function is described in the Kemp et al. SOM. We should have included reference to this to aid the reader. Calibration interval used was post- 1100 AD, see below for motivation.

For the present reply we undertook as requested to use the *entire interval* for calibration. This required a further downweighting of the observations in the likelihood function, and we conservatively tried a factor $2\times$. See figures, Fig.1, K11 model, Fig.2, G10*. We see a strong tendency to suppress the 20th Century upswing. For G10*, the a coefficient even becomes slightly negative. Further downweighting had little effect. Clearly this is not the way to go.

Next, we undertook to change the prior used for a_1. Our default prior was $U\,[0.01, 0.5]$ cm/yr/K, which puts quite a bit of likelihood at large values values of a_1. As pointed out earlier, our fit constrains the product $a_1 T_{0,0}$ but not a_1 or $T_{0,0}$ individually. This means that all combinations, from small a_1 with large negative $T_{0,0}$ to large a_1 with small negative $T_{0,0}$, enter the ensemble. We attempted alternatively to fix the a_1 prior more tightly, to $U\,[0.01, 0.03]$ cm/yr/K, with $T_{0,0}$ implied of -1 to -3 K. Doing this, *and no downweighting*, made it possible to obtain a much better looking fit over the whole period, see Fig.3 below.

Doing the fit only to the data post- 1100 AD however (Fig.4) showed that the tendency of the model to turn down sharply is still there, although now the data is inside the two-sigma error band. But this early data causes stress in the whole-period fit (Fig.3), as witnessed by a small but noticable compression of the 20th Century upswing. So the fit is better, but not good.

We also tried how much correction to the Mann et al. 2008 temperature data pre-1100 AD it would now take to make the downward turn go away, and found ~0.15 K.

(Personal note (MV): in developing the Kemp et al. manuscript, I worked with the full data (and strongly downweighted observations) for quite some time trying many things without obtaining satisfactory fits; the effect of excluding pre- 1100 AD data was a real find. The second such find came from the suggestion by SR to test the effect of a small ($0.2°$C) temperature bias in explaining the pre- 1100 AD discrepancy.)

You are comparing G10 and K11 while ignoring all other semi empirical models. We consider the G10 model to be superceded by our Jevrejeva et al. 2009 model, and the analysis therefore feels a bit dated. Especially as you try to use your conclusions to talk about the long-term projections of Jevrejeva et al. 2012, and Schaeffer et al. 2012. (Remark: counter-intuitively Jevrejeva et al. 2009 is actually more recent than Grinsted et al. 2010.).

Above, I have suggested a few changes to the analysis:

- Allowing for a constant residual GIA in the G10 model.
- Apply a recent base-line before calculating misfits and likelihood.
- More realistic uncertainties on the NCRSL record as a proxy for global sea level.
- Flat priors in log-space for positive / multiplicative parameters.
- Use the full period of overlap for calibration.

How robust are the K11 long term projections to these changes? Do these changes result in greater or smaller long term K11-projections? Does it bring G10 and K11 closer together or further apart?

See above.

DETAILED COMMENTS:

P3552 Abstract: It is stated that K11 gives markedly better fits. This is unsurprising given that the K11 model is the G10 model with a few additional terms. Obviously it will mimic a wider range of behaviors and fit a greater range of curves. I do not need a study to show me that. I also question the marked improvement. The K11 model seems to result in a huge misfit pre-1000AD. How does the sensitivity of the perpetual term change if you calibrate over the full proxy record?

The explanation of why the full record was not used is given in the manuscript (section 5, "The discrepancy before 1100 AD"). In short: a small systematic offset in the proxy-reconstructed temperatures for this period can explain all of the discrepancy. Such an offset is entirely plausible given that this early period is based on a small number of temperature proxies of geographically poor coverage. Therefore, we have less than full confidence in the fitness for purpose of this early temperature data, and do not wish to use it. This limitation of the early proxy temperature data is well known, see for example Mann et al. (2008). This problem once again highlights the nature of proxy data and the necessity of assuming it to be globally representative. No single proxy dataset for temperature or sea level is the global mean but development of semi-empirical models with longer datasets require their use.

Of course one cannot be certain that such a temperature offset is the cause of the discrepancy seen, but it is the most parsimonious explanation.

P3552 Abstract: It is stated that there is disagreement 2300-2500 AD. I would be very surprised if they don't agree within their very considerable uncertainties. Have you examined whether the confidence intervals overlap? Long term projections to AD2300 is done in Schaeffer et al. (2012) using the K11 model. This can be compared to the Jevrejeva et al. (2012) long term projections using a model similar to G10. For RCP4.5 at 2300 Jevrejeva et al. gets 70-287cm, whereas Schaeffer et al. (2012) projects a central value of ~350cm with an estimated uncertainty interval ranging from about

60% to 150% of that. I.e. although they are different, then their confidence intervals overlap by almost 80cm.

Yes, it is satisfying to see that the confidence intervals overlap. The projections are not worlds apart, and likely neither is very wrong. Perhaps we should have made this more explicit.

(note: It is unclear if the uncertainty reported by Schaeffer et al. includes the uncertainties both due to uncertain temperature projections, and from the K11 parameter uncertainty.)

(Schaeffer, personal communication) yes, it includes both uncertainties. Ensemble size for K11 fit was 1700, for RCP runs, 600.

You calibrate the model on NC RSL 1000-2000AD (in one of the figures). In the hindcast period we see that as soon as the model is let free, then it diverges from NC RSL with a too high rate. After 300 years this amounts to ∼40 cm. It seems reasonable to expect that K11 (and Schaeffer et al. projections) will be biased high by this amount.

But, as Figure 3 shows, the K11 model calibrated to only 300 years of NC data remains within one sigma for 700 years more, and within two sigma for 1200 years more. But yes, biases are to be expected. However, based on this we don't expect them to be large compared to the formal uncertainties from the Bayesian solution.

P3570 Line 1-2: I strongly disagree with them being "compatible". Please calculate a p-value. Note how it is only the corner of the green box (fig 4) which overlaps and the corner of the green box is not nearly as probable as the center bottom. The dating error is independent of the RSL error and therefore it would be much more appropriate to plot the New Zealand uncertainty interval as a skewed circle (which would be smaller than your green box). If you did that then it would be apparent that only an extreme coincidence in the two errors would allow the New Zealand data to be 'compatible'. When the confidence intervals only just barely touch then the probability is very small in

the overlap (you are multiplying two small probabilities). Consider this artificial example where there is no dating uncertainty where we have two measurements: [x1 = -20 ± 10cm ; and x2 = 5 ± 20cm (2 sigma uncertainties)]. The difference would be x1-x2 = -25 ± 22 and as that does not span zero, then these two measurements are incompatible.

The blue and green boxes give one-sigma uncertainties in sea level. Dating uncertainties are more ambiguous, but we don't actually use them. We apologize for the omission.

P3554 l21- p3555, line 15: This is an example of the one sided presentation that permeates the entire manuscript. The G10 model is called 'physically wrong' and it is stated that it has 'conceptual problems'. However, the additional term in the K11 model does not address these problems. Indeed, the perpetual response term in the K11 ensures that the K11 model simply has no equilibrium at all. It will result in an infinite sea level rise for constant present day temperature! If that is not unphysical then I do not know what is. I realize that both models are approximations to reality and obviously cannot capture all physical processes, -the perpetual term can possibly still be a useful if applied carefully to short time intervals. I highlight this 'unphysical' behavior of the K11 model only to show that the presentation is clearly one sided and simply unfair.

Agreed, we stand corrected.

P3555 L3-L15: Concerning the steric response. The G10 model (and the Jevrejeva et al. 2009 model) is equivalent to a one-box model, and we know that such models are able to mimic the full steric response from full AOGCMs on century time scales reasonably well which tend to show a roughly exponential equilibration to a step change (e.g. IPCC TAR fig 11.15). I.e. it has sufficient complexity to capture the key processes as best we know them. I feel that the authors are making an overly literal interpretation of Seq as the true long-term equilibrium rise. The G10 model is obviously an approximation and choosing a, b, and tau in the G10 model allows you to have a model that results in the correct present day rate.

Yes, G10 (and J09) is not at all a bad model.

P 3555 L6-9: Here the authors talk about the non linear response of ice sheets under extreme forcing scenarios. I agree that this cannot be modeled by the linear G10 model. However, the K11 model is also linear and will also not be able to such non-linear processes. Please note that in G10 we motivate our formulation as a linearization for small changes in forcing. I approve the presentation of such caveats, but at the presentation is highly imbalanced. All caveats come across as deficiencies of the G10 model although they are clearly shared by the K11 model. One more non-linear caveat is the finite reservoir of small glaciers.

P3558 l22. I would add geoid effects to these sentences.

Isn't this item 2, "The change in the Earth's gravity field when land-based ice melts"?

P3557 l25: It is not clear what calibration interval is chosen. It is obviously not the full overlap between Mann08 and the Kemp et al. NC RSL record as the K11 model deviates strongly in the early part of the proxy record. Please clarify and explain why you do not use the full record for calibrating the K11 model. Please also document the exact formulation of the entire likelihood function.

The explanation of why the full record was not used is given in the manuscript (section 5, "The discrepancy before 1100 AD"). The likelihood function used was the same as used in Kemp et al. (2011) and documented in the supplementary information there – we could have mentioned this more clearly in the present manuscript.

P3560 Line 7: Here the 'agreement' between the GIA corrected North Carolina proxy record and the J08 GSL reconstruction is highlighted. Yes, there is agreement within uncertainties, but tide gauge records from the region show that the applied subsidence correction is not sufficient to explain the difference with the rate of global sea level rise. If, as the authors argue, local RSL=GSL+GIA then we can calculate the GIA from the trend of RSL-GSL at the tide gauges. This is shown in review figure 2 for stations on

the east coast as a function of distance from Churchill, Canada. This is similar to suppl. figure DR3 from Engelhart et al. (2009) but corrected for J08 GSL. From review figure 2 it is clear that the Kemp et al. (2011) GIA correction falls well below the subsidence rates within the peripheral bulge. Using other GSL reconstructions with smaller 20th century trends will just make matters worse. This demonstrates that either the GIA correction is too small (as argued by Grinsted et al. 2011), or that NCRSL deviates substantially from GSL. In either case it cannot be used to calibrate semi-empirical models. One of the co-authors even acknowledged as much. Quote:

> - '*If, as you say in your email, it was to point out that "it is premature to calibrate a global semi-empirical model to this single proxy record" then I would have been very supportive...indeed it's the basis of our current research and many others within the sea-level research community.*' – Benjamin Horton (10. Dec 2011).

It is also clear that it cannot be used for formal validation or formal model selection. Another critical aspect is that there is not agreement between millennial scale paleo sea level proxies. Kemp et al. 2011 figure 3 shows a large disagreement with non us records such as Iceland, Israel, Southern Cook Islands. The same figure also shows that virtually all other shown proxy records have higher RSL than the Kemp reconstruction at 1000AD. The early NZ data (fig 4) is also above the NC record. To me this is suggestive of a too low subsidence correction. All other lines of evidence (20th century tide gauge rates, GPS data, and model results) also suggests larger subsidence corrections (Grinsted et al. 2011; review fig. 2). In your reply to Grinsted et al. (2012) you argue why you prefer this outlier estimate, but it remains an outlier and there is a considerable spread among estimates. In light of this it is clear that your confidence intervals are overly optimistic.

Our motivation for and belief in using geological data to estimate the contribution made by GIA to subsidence to the relative sea level reconstruction is presented in the original

Kemp et al. (2011) paper, our published response to Dr. Grinsted's comment and a short section was included in this manuscript. Our position is unchanged and it is evident that Dr. Grinsted's reservations are every bit as strong as our conviction that this is currently the best way to make this correction. Briefly, GIA models make the same assumption about eustatic change, but do not include any tectonic component to land level change (which geological data inherently account for) and have a demonstrated systematic misfit to RSL data (see for example Engelhart et al. 2009). GPS instruments do not yet have the necessary record length to provide a robust GIA estimate and do not actually measure GIA but the net effect of all land level motion. To call our estimate of GIA "an outlier" is simply wrong and deliberately misleading and displays a refusal to acknowledge the considerable limitations of other approaches. It is naive to include "20th century tide gauges" as a means to estimate the rate of GIA, these records have been correcting using a global GIA model to begin with and so using a rate to fit the tide gauges would be no different to using a GIA model. Indeed the problems with GIA models could be invoked to question the accuracy of estimates of 20th century sea-level rise from tide gauges.

The differences among sea-level records return once again to the nature of reconstructions developed by different researchers using different proxies, different dating methods and different approaches to estimating GIA. We retain our belief in the reliability of the North Carolina record.

P3559 L24- L26: I agree that the rate that you apply is consistent with Engelhart et al.'s (2009) estimates for neighboring regions. This is however, unimpressive as the subsidence correction is essentially based on Engelhart.

P3559 L 26-27: Yes the K11 model is able to accommodate any GIA rate without any loss of fit. But we are not just trying to fit a curve. Is it really an advantage that the model conflates a non-climatic local signal with a climate induced sea level change? It is important that you allow for this GIA uncertainty in the sea level observations when calibrating the model. You can do it through a C-matrix as in G10, or you can amend

the semi-empirical model to include a GIA term.

The reason we pointed out that the K11 model is able to absorb (to first order) any GIA value is to make clear that we are not motivated to oppose a very different GIA value if sound arguments in favour of such a different value are brought forward. The core results and conclusions of Kemp et al. (2011) would not change much, although obviously the numbers would change and the graphs would have to be re-drawn. Also the earlier Comment to Kemp et al. (Grinsted et al., 2011) pointed this out, and we agree.)

P3559 L14-17. Please quantify what is meant by "did not change appreciably". I take it to mean that abs(dGIA/dt) Âż abs(dGSL/dt). That is probably a reasonable assumption in locations where this method is typically applied (i.e. where there is an appreciable GIA response). So, I agree that dRSL/dt in such cases is a reasonable approximation to dGIA/dt. However, that does not mean that the residual GIA is neglible for your application. Specifically I dispute the narrow uncertainty range (±0.1mm/yr) on the GIA you derive from this approach (See also comments in Grinsted et al. 2011).

Yes, this text needs to be more quantitative. We still don't agree on the latter point.

P3559 L14-17. Even if global ocean volume did not change, then there could still be considerable deviations between local sea level and global sea level. Relatively small changes in circulation could result in a large local sea level response. How can you be sure that is not an issue when there is substantial disagreement between proxy records from around the world? (as shown by Kemp et al. 2011).

Once again this uncertainty is the nature of proxy data and it must be accepted if semi-empirical models are to be calibrated on time scales longer than the instrumental period. It affects the temperature proxy reconstructions too and we should not forget that it also affects the older parts of instrumental records which are hampered by limited spatial coverage. One of the most beneficial aspects of salt-marsh reconstructions is that salt-marsh ecology and sampling inherently remove short-lived variation and the

North Carolina record delivers multi-centennial changes in sea-level. As stated in Kemp et al. (2011) deviations of sea level from the global mean can only grow so large. We expect that there is variability around the mean instead. All of the proxy records have limitations which could be discussed.

P3560 L4: You are too optimistic regarding the proxy uncertainties. I get more from the GIA uncertainty alone, and that does not even allow for any differences between local and global sea level. Compare New Zealand @ 400AD with NC RSL (fig4), or look at all the other records that Kemp et al. (2011) shows. Is ±10cm really realistic, when there is not agreement between proxies from around the world? It is not convincing to me.

GIA uncertainty is treated by us as separate from the rest of the uncertainty budget. Yes, some of the discrepancies seen in Kemp et al. Fig. 3 could be from wrong GIA values.

P3560 L7: In order to arrive at as low a value as 24cm then you have to force the straight line fit to go through 0 at time of collection. Simpler statistical approaches would arrive at slightly higher values. If this is such an important constraint to ensure marginal agreement with tide gauge GSL, then you should also apply this constraint to the model fits. I.e. use a recent baseline in all plots to ensure that all models go through this point. The likelihood function should be based on the misfit when both model and data both use a recent baseline.

It appears that we do things differently. Note that we have a model parameter H_0, an integration constant for sea level, resulting in a different variance structure.

P3561: Nobody really believes the 'perpetual' term to be truly perpetual. Another reason for the large disagreement could be that the K11 model simply cannot be applied to so long time scales because it does not capture the natural equilibration that would occur over long time scales. That would mean that the Schaeffer et al. 2012 projections would tend to be biased high (see section 7.7).

This is not impossible. But making the perpetual time scale finite would mean introducing even more tuning knobs still, with little in the way of data to set them. We did not want to go there.

Also consider this: $T_{0,0}$ is negative, and so would a $T_{0,0}(t)$ for a finite time scale be. And it would increase over time relaxating toward $T(t)$, making the pre- 1100 AD divergence worse instead of better. So, making the perennial time scale finite won't even be sufficient on its own.

See also our reply to the Robert Kopp review.

P3564 L15: You examine only two GIA cases (figs 1&2), none of which necessarily are correct. The primary argument of the Grinsted et al. (2011) comment to the Kemp NC RSL paper was that the reported GIA uncertainties are too optimistic, and that the particular subsidence rate applied is an outlier (all other methods result in higher subsidence rates - e.g. review fig.2). We do not argue that the Peltier GIA is correct, we only use it as it is in the middle of the range of estimates (actually lower middle). It is therefore a strawman to show that G10 does not work for Peltiers GIA. A fair comparison would instead allow for a range of different GIA's. Instead you should allow for a small residual GIA into the Grinsted et al. model, before attempting to fit. Seq=a*T+b; dS/dt=(Seq-S)/tau; NCRSL=S+rGIA*t. I can get a much better fit to NCRSL with this G10+rGIA than with K11 (both forced by Mann08).

C1418 figure 1 shows the G10+rGIA+Mann08 model with these parameters [a=0.7;b=-0.11; tau=100;rGIA=0.0005]. It gives results that are rather similar to the Kemp NC RSL record.

We commend the reviewer for engaging to the point of performing his own calculations. Yes, the agreement looks not bad, and we obtain similar results from these assumptions (see, e.g. Fig.5 below). However, there are problems with this result:

1. The very short τ value, confirming what we found, that the G10 model favours

short τ values. While these are not impossible, they exacerbate the discrepancy with the much larger sea-level variations found in the geological record, as discussed in the manuscript

2. The very large correction to the GIA value, 0.5 mm/yr (resulting in a value of 1.5 mm/yr). Back to the year zero, this correction corresponds to a whole metre. We do not accept that either the proxy observations are this inaccurate, or that stationary sea-surface topography has changed by this much, at the North Carolina site without detection in sea level records of from other forms of geomorphic evidence

3. There is still stress in this full-period fit, visible, e.g., as a compression of the 20th Century upswing.

Fig 1: I can get much closer to agreement between the G10 model and NCRSL with 1.3mm/yr subsidence correction than the plots you show (with G10_tau=100). This makes me question whether you give the G10 model a fair shot. What is the calibration interval, and how was that chosen? How does the likelihood function look?

See above.

P3555 equation 4. The concept of a purely mathematical moving reference temperature T0 is bizarre. And the use of T0, T0,0 and T0(start_of_integration) is highly confusing. I would much prefer that you re-formulated the model along these lines: dS/dt=dPerpetual/dt + dMedium/dt + dImmediate /dt where each term has its own differential equation. It would make the physical motivation for the model much more apparent and it would make it much easier for the reader to compare to the G10 model. This comment is purely about presentation.

Point taken ;-)

However, we do not want to deviate from the formalism already presented in Kemp et al. and also used in Schaeffer et al.

P3565 line 19: An alternative explanation is a residual GIA term. I can get reasonably good fits when I allow for a residual GIA term. On the other hand, my exploration of the K11 parameter space reveals that it is impossible to fit the early part of the record without sacrificing the fit in the most recent centuries. Perhaps this is what led the authors to restrict their calibration interval.

Actually, no. Yes, it drew attention to these early data. But we have a plausible and parsimonious hypothesis for why these data should not be used.

P3569 Section 7.6: I appreciate that the authors have tried to confront the problem that NC RSL is not global sea level by using proxy records from another location on earth. However, that does not solve the issue. This tests against two records – neither of which is global sea level. And as the New Zealand data demonstrates then there is still large disagreement between the proxies on millennial timescales. Either the applied subsidence correction is wrong, or the records contain some local signals. Either way it is premature to use these proxy records to calibrate global sea level models.

At present there is no means to resolve this issue in a fashion that Dr. Grinsted would find acceptable. We return again to his fundamental concerns surrounding proxy reconstructions weighed against the necessity of using them to calibrate models using data older than the instrumental record. Despite this the G10 publication selects a data set from fish tanks in the Mediterranean Sea as being representative of the global mean.

The NC proxy data is unique, in its precision (due to local circumstances), the volume of data, and its redundancies. Adding many different proxy reconstructions like those in Figure 3 of Kemp et al. is not a solution. Note that disagreements between proxies have many causes (e.g., methodologies), not just differences between local and global sea level.

We chose the Tasmania/New Zealand data set because it is good quality, generated using an approach and techniques very similar to the North Carolina reconstruction and is spatially independent of North Carolina to provide a record from outside the Atlantic Ocean basin and the influence of Laurentide GIA.

P3570 section 7.7: If you want to argue that K11 is better than Jevrejeva et al. 2009, then you should write a paper comparing these models. Note that J09 does not use temperature as forcing and the shape of its response would not necessarily be the same as the G10 model.

We hope that such a paper gets written. What we did in the current manuscript was replicating the typical appearance of the G10 sea-level curve, and the associated short τ value, using as input both different temperature and different sea-level data sets. The only commonalities with the original G10 were 1) use of the G10 semi-empirical formulation, and 2) use of sea-level data post- 1700 AD only. We find this convincing, and are surprised that the reviewer does not.

P3570 section 7.7: The K11 model diverges quickly from the NC RSL record with a too high rate pre-1100 AD. In order to gauge how quick it deviates from the curve then it would be good to know the calibration interval you chose for the K11 model. At the moment I can only find this in the figure caption, and no motivation is given for restricting the calibration. Has this been chosen to present K11 in a favourable light?

The motivation is discussed rather extensively in the text, and was already discussed at some length in the Kemp et al. Supplementary On-line Material.

P3570 line 13: You state Schaeffer has a much slower equilibration, but that is not correct as the K11 model does not have an equilibrium at all. I think "deceleration" would be more accurate.

OK ;-)

P3566 Line 17-20: I find this argument unconvincing and hard to read. Please phrase these sentences more carefully.

Agreed, thanks.

P3567 Line 9-P3567 L4: I acknowledge the Milankovitch issues that there is with using previous interglacial directly as analogies. But I strongly dispute the argument that the situation gets any better when you go to Mega-year timescales. Both sea level and temperature proxies are much more uncertain the further you go back in time; Milankovitch is also active on these long time scales; And you have to keep non-climatic processes in mind (e.g. plate tectonics).

P3567 Line 9: I feel you make an overly literal interpretation of the G10 a-parameter. G10 is an approximation and we do not imagine that this simple model will be sufficient for the far future where we may reach equilibrium. At some point the model will need to include a long-term (but not perpetual) component. I doubt that we have sufficient information to constrain all these additional parameters. In particular, I am worried that you are over-confident in the NC RSL record and that this in turn leads to over-fitting of the K11 model.

We don't think we are over-confident. There are important caveats, but those apply as well to the G10 approach – and in fact to the state of the art independent of model. And any plausibly long-term component big enough to make a difference, would show up on the millennial time scale too.

P3569 L28: Typo. G11 should probably be G10.

Yes, thanks.

P3570 L23-L25: Actually this term should correspond to time scales that equilibrate with time constants longer than the time interval the model is applied to. Otherwise you need to take the deceleration into account. I.e. it is wrong to compare to tau here.

Yes, we can see that it can be misread that way, and needs to be reformulated.

P3570 L25-26: Sea level has been equilibrating since the de-glaciation. It seems counter-intuitive that a perpetual non-equilibrating term would capture that. You can capture the slow down by strategically moving the reference temperature. However this is essentially equivalent to having a finite response time.

P3571 L4-7: I find that the perpetual term makes it impossible to fit the pre-1000AD data without sacrificing the fit post-1800AD. So, the statement is only correct if you have a blind spot that restricts the interval that you are looking at.

There are reasons for eliminating the pre-1100 AD data, see above.

P3571 L7-L14: Again I feel that this is an over-interpretation of Seq in the model, and that these sentences don't allow for the full range of response times explored by the Jevrejeva et al. 2012 model. If you use the G10 model for multi-century projections, then you may get too low estimates unless you take the models with a long response time. - Just as the non-equilibrating response of the K11 model will tend to be biased high for long term projections.

P3571 L15: Agreed.

Figure 1: I suggest removing this figure and instead allowing for a constant residual GIA term in the G10 model in figure 2.

We do not wish to do that for reasons explained.

Figure 1: Please remove blue line. I believe you when you say that you can reproduce the results.

Hmm. But the agreement between Bayesian median and simple integration is informative.

Figure 2: This figure shows that the additional terms in VR09 probably led to over fitting as it appears to be worse than R07.

Both curves were obtained by "tuning" (manually!) the T_0 value as described in Kemp et al. We copy these curves as such. Due to this manual tuning element, no meaningful quality-of-fit comparison between them can be done. Apart from both models being not designed for, and unfit for, this time scale, even after T_0 tuning.

Figure 3: Please remove "Mann integrated" to simplify figure.

Good idea.

Table 1 & figs 1& 2: Where does the K11, tau=4000 yr come from?

4000 is a "surrogate" for ∞ in numerical calculations. We verified that making τ larger still causes only small changes and doesn't make any qualitative difference.

Table 1 & figs: I think it would simplify the discussion and all figures if you could focus exclusively on Mann et al. 2008. The Moberg experiments could be sent to the SI (except for the original G10 moberg/jones hind casts).

Yes, good idea.

Figure 2: This figure does not say much when it is based on manual tuning. Perhaps I would have made another more flattering manual tuning of the G10 parameters.

Instead of manual tuning we went to using a least-squares fit (see our reply to Kopp comment), with little change.

Figure 2: Why do you restrict the calibration to post-1000 AD? Why is tau fixed at 4000?

4000 is a surrogate for ∞ in numerical calculations, see above.

Figure 3: Here a post-1100AD calibration interval is used, in figure 2 a post-1000AD interval is used. Why is it different?

Clerical error, thanks. Post- 1100 AD is used.

Figure 3: The red, black, and blue is the primary lines in this figure. The other two lines

make it cluttered. It might help to de-emphasize them with dashes.

We tried and it does help, thanks.

Figure 3: Please do not hide the G10 and K11 lines for the recent period. (Choose a blue color for K11)

We lifted them to foreground and yes, it looks clearer. Thanks.

Figure 1,2,3: I think it would be good to have the PDFs with shared x-units also share the same x-limits. It might mean that you have to have a logarithmic x-scale.

In principle a good idea, but tricky.

Figure legends: Please find a uniform way to name the different hind-casts/ forcing combinations. For example in figure 3 you write "Mann et al. 2008 Bayesian", but it does not say that it is the G10 model. I think a legend like "G10 w. Mann08" would be easier to understand. Figure 2: You can remove the text "hind-cast" in the legend entries.

We followed up on this, see our reponse to Kopp comment, making the legend entries more uniform. Removal of "hindcast" text was a good idea.

Figs 1-3: Please consistently use either one or two sigma intervals in all figures (and text).

Good idea in principle, but we wish to keep one-sigma in Figures 1 and 2 as there, the colour bands are only background info. Elsewhere one + two sigma is informative.

Figs 1-3: It would help the reader if a consistent color scheme was chosen. E.g. all K11 plots could be blue tones, where as the G10 could use black (as in fig 3). Solid lines for those that you consider most important for the discussion and dashes for those that are secondary.

This is not so simple. We use grey vs. blue to distinguish Bayesian solutions, or parts

of solutions, in the same plot.

Figure 4: The green box (@400AD) is the same height on both left and right sides. Presumably the reason the green box is skewed is due to a vertical adjustment. If there is an uncertainty in this correction then the box should be higher on the 'older' side of the box.

The green box is skewed due to applying the GIA correction. The box does not contain the GIA uncertainty, which is treated separately (this has been our practice throughout, as GIA uncertainty is not random in the same way that other sea-level uncertainties are).

Figure 4f and g: Are you sure that the markov chain has converged?

We used a brute-force Monte-Carlo approach. Yes, especially figure 4g doesn't look very nice, perhaps due to the generally poor fit producing very low likelihoods, together with our limited ensemble size (25,000). Tried 250,000, no real improvement.

REFERENCES:

Jevrejeva, S., Moore, J. C. & Grinsted, A. (2012). Sea level projections to AD2500 with a new generation of climate change scenarios. Glob. Planet. Change 80-81, 14–20 doi:10.1016/j.gloplacha.2011.09.006

Schaeffer et al. 2012, Long-term sea-level rise implied by $1.5°$C and $2°$ C warming levels. Nature Climate Change. doi:10.1038/nclimate1584

Jevrejeva, S., A. Grinsted, and J. C. Moore (2009), Anthropogenic forcing dominates sea level rise since 1850, Geophys. Res. Lett., 36, L20706, doi:10.1029/2009GL040216.

Grinsted A, Jevrejeva S, Moore JC (2011) Comment on the subsidence adjustment applied to the Kemp et al. proxy of North Carolina relative sea level. Proc Natl Acad Sci USA 108:E781–E782.

Engelhart, S. E., Horton, B. P., Douglas, B. C., Peltier, W. R., and Tornqvist, T. E. (2009): Spatial variability of late Holocene and 20th cen- tury sea-level rise along the Atlantic coast of the United States, Geology, 37, 1115–1118, doi:10.1130/G30360A.1, 2009. 3559 (suppl.info can be found at
ftp://rock.geosociety.org/pub/reposit/2009/2009276.pdf)

Tarantola (2004), Inverse Problem Theory, SIAM, ISBN 0898715725, p161
(http://www.ipgp.fr/∼tarantola/Files/Professional/Books/InverseProblemTheory.pdf)

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F. (2008): Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, P. Natl. Acad. Sci., 105, 13252–13257.

Kopp, R.: Interactive comment on "On the differences between two semi-empirical sea-level models for the last two millennia" by M. Vermeer et al., Clim. Past Discuss. 8, C1310-1316, 2012.

Interactive comment on Clim. Past Discuss., 8, 3551, 2012.

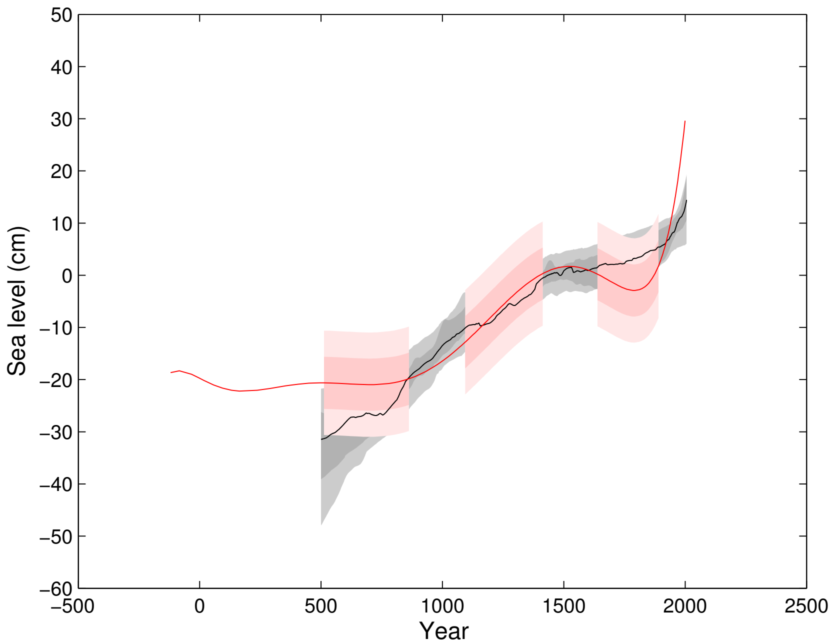
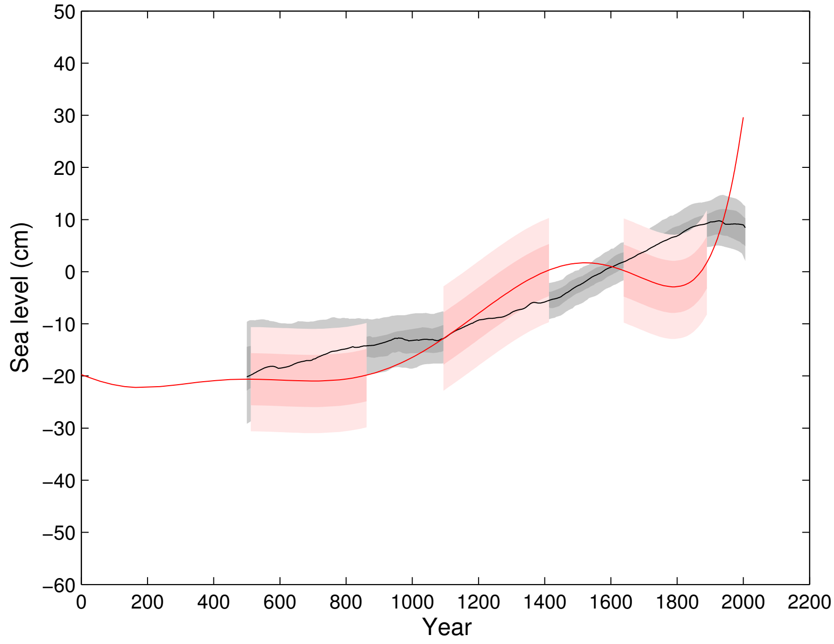
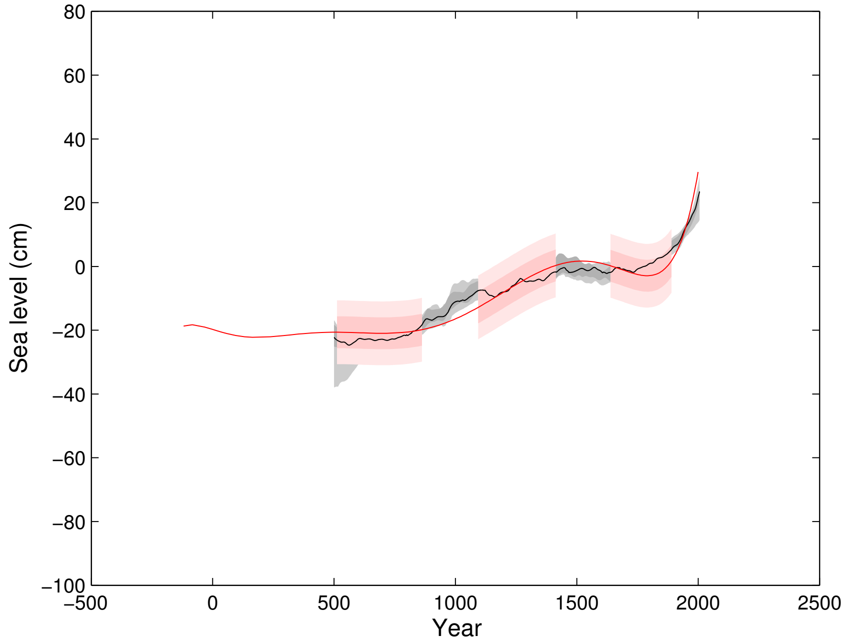


Fig. 1.



Sea level (cm)
Year
50
40
30
20
10
0
−10
−20
−30
−40
−50
−60
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200

Fig. 2.



Sea level (cm)
Year
80
60
40
20
0
−20
−40
−60
−80
−100
−500
0
500
1000
1500
2000
2500

Fig. 3.

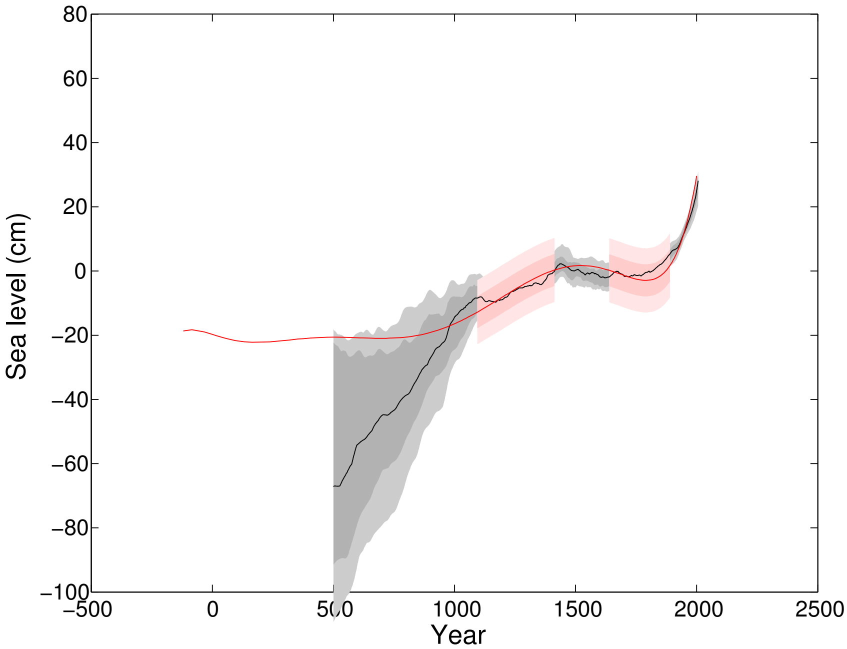

80
60
40
20
0
−20
−40
−60
−80
−100
Sea level (cm)
−500
0
500
1000
1500
2000
2500
Year

Fig. 4.

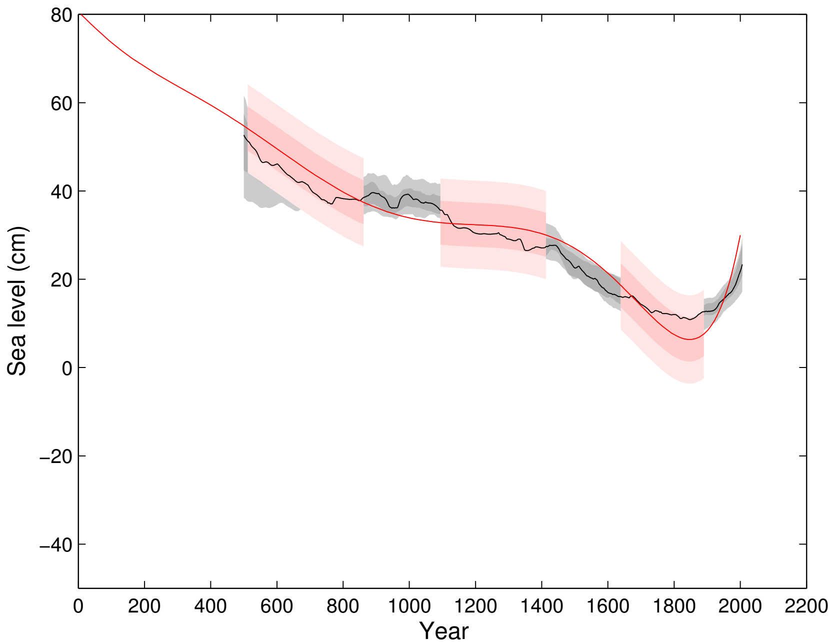

80
60
40
20
0
−20
−40
Sea level (cm)
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
Year

Fig. 5.