

Interactive comment on “Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium” by A. Hind et al.

G. Hegerl (Referee)

gabi.hegerl@ed.ac.uk

Received and published: 12 March 2012

This ms sets out a much needed statistical framework for reconstructing and analyzing reconstructions, and for determining if externally forced signals are present in the reconstructions, and then tests the proposed methods within a series of climate model simulations (perfect model test – and pseudoproxy approach). I like the idea pursued in the paper, but have still quite a few questions / suggestions

1) The paper is a very long read. Maybe it can be shortened a bit by pulling out some more stuff into an appendix. Also, numbering the equations would help readers to jump around in the paper. Given that such a proposed method would be useful for a

C114

wide range of work, it would be good to make sure this is read widely. The present format is not as accessible as it could be 2) Some assumptions seem too stringent, I am particularly thinking of white noise temporally– this is particularly a problem for delta (the noise in the model) and eta (the internal variability in the data) which is going to be spatially and temporally correlated and not limited to particular timescales (Hasselmann, 1976; note that even if you cant tell statistically that temporal correlation is present beyond 20 years, this doesnt mean it isnt there if physics tells you it will be there). Also, neither climate model runs nor data are strictly red, they have more complex temporal behaviour. Therefore I find it more helpful to use the climate model control run variability rather than a white or red noise model, as the latter may capture the climate variability and its different timescales better (not saying its perfect, but less imperfect than a red noise model). You do that for one of your statistics. 3) The U_r and U_t metrics proposed are useful and interesting, but I wonder what their relationship is with a very closely related framework pursued in fingerprint detection. To me, fingerprinting could be easily extended to proxy based data, and uses an underlying similar understanding of signal and noise to that pursued here. To me, fingerprinting has two advantages, which I think makes it more useful for estimating external signals and determining the causes of forcing: Fingerprinting can straightforwardly deal with several external influences simultaneously, and determine which of these have left imprints on the data that are identifying from residual variability AND each other. This problem of degeneracy is implicitly mentioned in the paper when discussing that the volcanic signal is more or less detectable given the size of the solar signal in the ensemble used. References are Allen and Tett, 1999; Hegerl et al., 2007a (for overview) b for last millennium application, and Allen and Stott, 2003, clim dyn 21 for the total least square method appropriate here. 4) The second advantage is that fingerprinting couches this as a regression / estimation problem, and hence the answer is not a ‘yes/no’ answer but the estimated amplitude also indicates if the signal is similar to, larger or smaller than the model signal. Hence, the complicated answer provided to the question which solar signal matches better with the data can be addressed simpler, particularly if as-

C115

suming that the main uncertainty is in the magnitude of the solar forcing, by estimating its magnitude from data. The hitch then is only to make sure that ensembles are large enough for the solar forcing in the run not to be swamped by the noise in the run. I worry that in the present framework, the solar signal similarity will be determined by ensemble size, possibly leading to the result that an Energy Balance Model Solar signal may have more skill (as there is no noise) than a full climate model solar signal that struggles against variability such as in E1. Therefore I recommend to at least discuss this alternative approach, pursued to some extent (but not in the full pseudo-proxy approach) by Hegerl et al., 2011 and 2007 and cross relate to where both are similar and different. The total least square approach for calibrating and detecting signals from Allen and Stott is philosophically similar to the statistical model 1, but in my view with fewer limitation. 5) The finding that limited spatial coverage is sufficient to estimate large-scale long-term signals is a really useful one, good to highlight. I find the same in my papers, and I found that reviewers tended to be sceptical about this, as the recognition that you don't need that many spatial points for decadal temperatures is not that widely known. 6) I dont think ranking models based on agreement is a good idea given forcing and proxy uncertainty – although you could shed light on this by finding which models are indistinguishable from each other.

Detail comments Introduction: I like the way the context is set out. Statistical model 2: this looks exactly like 1 just without the forcing term – is a separate model needed, wouldnt just specifying a zero forcing be clearer? p. 275, top: This is not the 'normal' calibration but is what is often called inverse regression (advocated by Christiansen, and is a simplified version of tls where noise on both regressor and regressand is considered) unless I am mistaken. p. 274: I am a bit confused here – it seems that there is a bias due to theta as its variance increases the difference – what am I missing? Also I am not sure if eqn. (2) (numbers on very few equations – distribute further) is fully explained, and errors in variables needs quotes as its a quite widespread approach. Also refer to papers having tried this in the literature. p. 277: this is only an ideal weight if the data are white – optimal fingerprinting allows for nonwhite noise by using

C116

the inverse covariance matrix here.

P, 284: U_T is hard to understand backwards as I have lost at this point what the c_i are (other readers might as well). Same for the U_R statistics. It should allow a lazy reader to see the equation and follow it back easily.

Figures: Figure 2 and 3 is nice and clear, but you should mention what E1 and E2 are for a reader who is checking this quickly. I like that you used the proxy network for figs 5. Figure 6 ff have no x-axis labels and its not immediately obvious what it is without ploughing into the paper. Note also for the coverage problem and the covariance matrix becoming hard to do for large number of regions: this can be easily overcome by spatial truncation to a better base space, eg EOFs.

Interactive comment on Clim. Past Discuss., 8, 263, 2012.

C117