A quality assessment of AR5 Climate Models and its implication in underestimation of localized Climate Change

Xavier García Casals (xgcasals@gmail.com)

12/12/2013

Submitted as Term Project for the 'Global Warming: The science of climate change' course from the University of Chicago

Introduction:

Climate models have been doing a great job, along the last 24 years since the publication of the first IPCC report, in setting the basis for the understanding of the Earth's climate system response to anthropogenic inference, and building up a consensus against the resistance of 'convenience denial'. From basic 'pencil and paper' models to today's 1-million code lines models, all of them have contributed to build up the evidence of climate change and its origin in GHG emissions.

Still, taking this fact by granted, there is still uncertainty on the level of climate change to be expected if we follow one or another emission path, especially on the regional and local level, where true impacts are going to take place . Climate models are used to quantify this, but perhaps a too conservative approach has been used when deriving the probabilistic distribution of climate change from the population of the different climate models, giving each model's results the same weight, without consideration to the quality of each model. In fact, straight information regarding the relative quality from the different models used in the IPCC reports is not available, preventing the capability to discriminate model's results beyond the general statistical presentation they are provide with. At the end of the day this means that the focus of attention goes to the mean or median of model's results, while this statistical indicator may be compensating the inputs from models with very different capability to predict climate change in a specific location.

A preliminary fast scanning of the comparison between historic data and AR5 model's predictions with historic forcing, in regions which are currently water stressed, or where water stress and drought severity indexes are expected to grow with climate change, gave me the impression that global climate models could be significantly underestimating[1] the climate change that we could expect in these sites, which are among the ones where the highest impacts from climate change could be expected. If we add to this the fact that current climate models still do not incorporate 'all' the physics that govern climate change (because of unknowns and computation capabilities), and that probably most of what is left out from them goes in the direction of intensifying climate change[2], we could come to the conclusion that perhaps the expected climate change reported up to now has been conservative.

Thesis:

The climate models used for the AR5 IPCC report have significantly different quality, as defined per their capability to reproduce local climate. Using these model's results without factoring in their differential quality can lead to significant deviations in local climate change predictions. For water stressed regions, where climate change impacts can be highest, these deviations seem to heavily underestimate local climate change, and therefore much higher impacts than predicted could unfold.

[1] In these regions we could expect the highest impacts of clouds and aerosols, two of the components that add highest uncertainty to current climate models.

[2] For instance additional feedback mechanisms, or increased speed response of what are thought slow feed back mechanisms.

Methodology and some results:

A site in Spain (Valladolid) has been taken to illustrate this point. Figure-1 presents historic temperature records and predictions from 11 AR5 climate models when forced with historical radiative forcing. As it may be seen, although there is a lot of noise (both from models and from short time scale climate effects), there seems to be a tendency to underestimate (in average) local temperature evolution, and very heavily in some climate models: What about if the 'best' quality climate models are the ones predicting the higher temperatures and the 'worst' quality models predicting the lowest? This would lead to an underestimate of local climate change that would go much beyond the one reflected by all the model's mean.

But certainly from all the 'noise' in Figure-1 it is not possible to extract more conclusions, so some steps are required to get information out from these signals. These steps provide the basis for having a methodological relative quality assessment[3] of the different climate models for predicting local climate.

First step is a calibration[4] to local conditions. Regionalization climate models do so by taking as boundary conditions for their runs the results from the GCMs. We are going to do so by adjusting the model's results to the average historic data in the period[5] from 1867 to 1898.

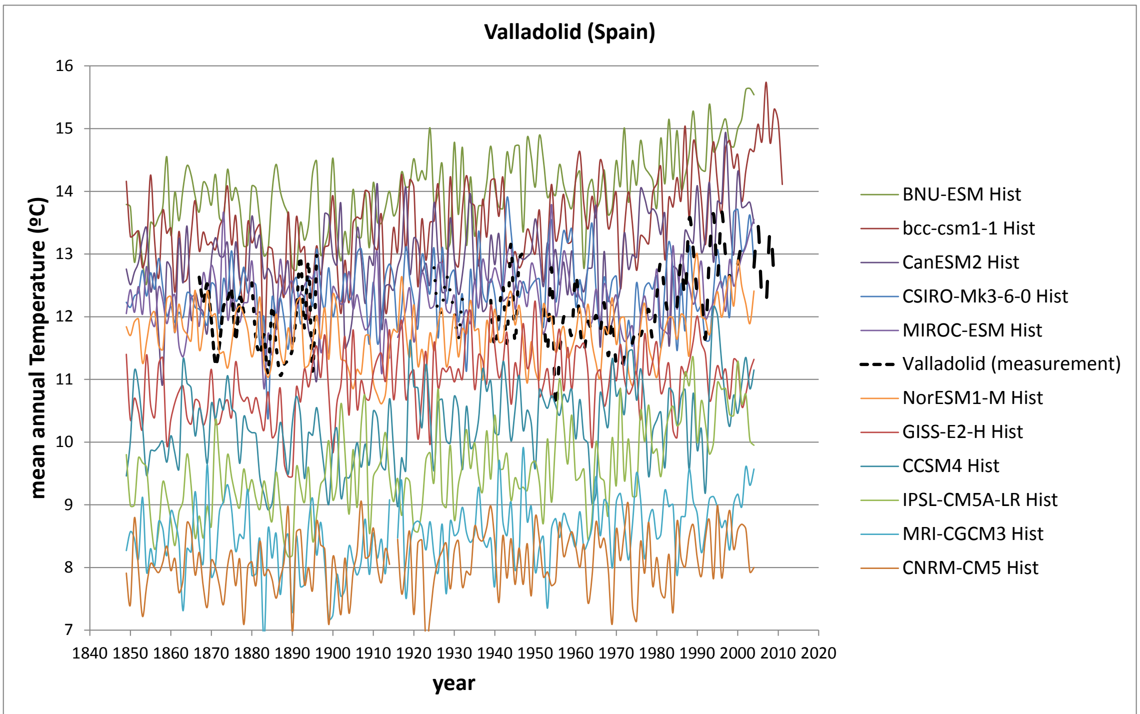

Figure-1: Measurement annual mean temperature in Valladolid, and AR5 climate model's reconstruction of the historic value driven by historic natural and anthropogenic climate forcing, accessed from the time series browser from the University of Chicago (meteorological station reference: #Kcg).

[3] The outcome of this quality assessment can be different for different sites, therefore I don't mean with the reported results to make any claim on absolute climate model's quality.

[4] In terms of temperature anomaly, this calibration is equivalent to the normalization process used to compare model's temperature anomalies.

[5] We took this period because of having continuous measurement data in the meteorological station and because is before the industrialization anthropogenic impact.

Second step will be filtering the signals from Figure-1 to distillate the main information they contain. A convenient way of filtering is by using regression polynomials that provide the best fit to the data. The degree of the polynomial is relevant because in general, the higher the degree, the more information it contains from the original signal[6]. Often the main focus goes only on temperatures, and at most in average gradients, but in order to evaluate the capability of a climate model to predict local climate change (its quality for local predictions), we are interested in the evolution of the precursors of change: temperature gradients and curvature. A model that is unable to properly capture the evolution of temperature gradients and curvature, has a questionable capability to predict future climate evolution (which is driven by these parameters), and therefore should be assigned a lower quality.

Figure-2 presents the gradient evolution from cubic fits to the historic measured and modeled data. As it may be seen the different models certainly have a very different capability to reproduce the measured evolution, and in fact some of them seem to be completely 'out-of-phase'. It is also worthwhile to point out from this figure that the gradient as today is lower in all the models than in the measured data, which again could point in the direction of underestimating climate change.

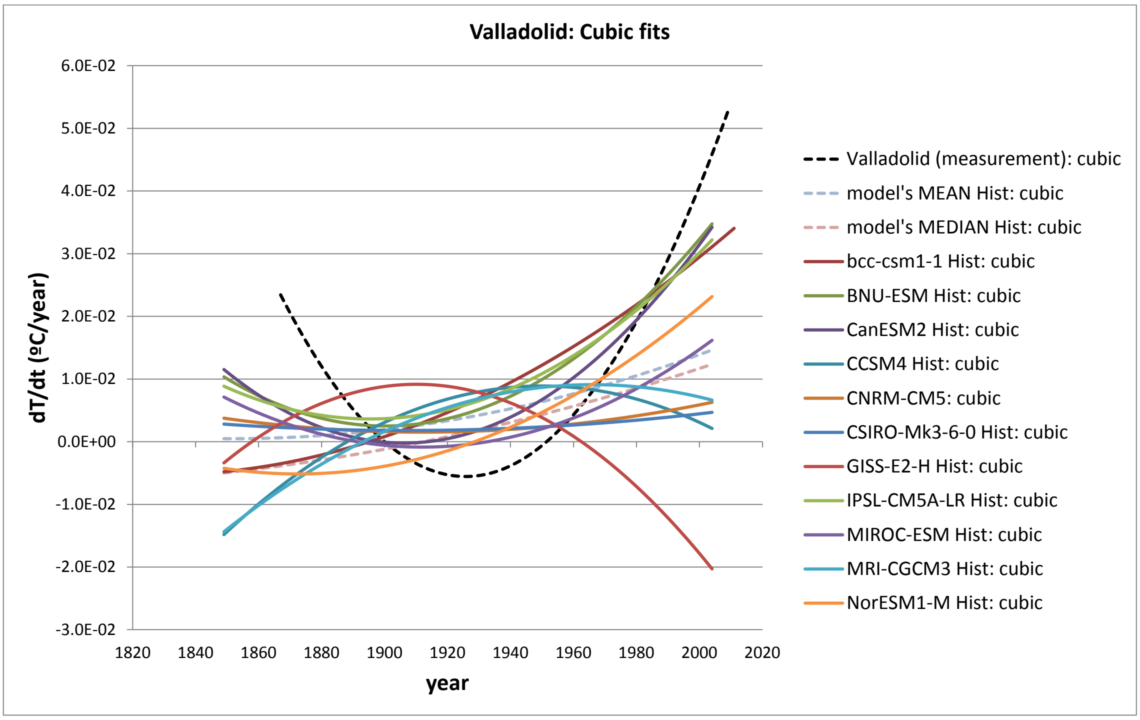

Figure-2: Comparison of the historic evolution of the mean annual temperature gradient (dT/dt) from third order regression polynomials of the measured data and each one of the 11 AR5 climate models historic output when fed with the historic total (natural plus anthropogenic) forcing. The regression polynomials associated to the model's mean and median are also shown.

[6] We are quite used to first order polynomials regression (linear regression). However, the information contained in such regressions is quite limited, and they do not even manage to capture the 'hockey stick' temperature evolution in the last years, which is central to the anthropogenic origin of climate change. Indeed, a linear regression has a constant gradient, and therefore has no information on gradient evolution, which is the driver of climate change.

In order to quantify the quality from the different models to reproduce historic climate, we computed the relative residuals from the calibrated temperature signal, the temperature gradient and the temperature curvature. Table-1 shows the results when using a fourth order polynomial to filter the signals. As we may see, residuals in temperature gradient and curvature are much higher than in calibrated temperature, and therefore certainly need to be taken into account when evaluating the relative model's quality to reproduce a local climate. The last column of this table shows a composite residual obtained by normalizing each residual with the highest value, and weighting the normalized values (35% for temperature, 40% for gradient and 25% for curvature) to get a compound residual, and is the criteria we used to rank the different climate model's quality for representing the local climate.

model	T-calibrated	dT/dt (gradient)	d^2T/dt^2 (curvature)	normalized weighted composite
MIROC-ESM	0.9%	40.7%	23.8%	24.8%
CanESM2	1.7%	48.2%	7.8%	32.1%
CSIRO-Mk3-6-0	1.1%	77.6%	45.0%	39.8%
NorESM1-M	1.0%	75.8%	71.6%	44.8%
CNRM-CM5	0.7%	88.0%	83.7%	46.7%
BNU-ESM	2.3%	73.6%	49.4%	53.2%
bcc-csm1-1	2.6%	75.6%	49.9%	57.9%
IPSL-CM5A-LR	3.1%	82.2%	47.0%	64.5%
CCSM4	1.5%	123.5%	107.8%	68.8%
MRI-CGCM3	1.3%	126.8%	116.4%	69.7%
GISS-E2-H	2.6%	169.3%	103.3%	90.9%

Table-1: Residuals for the calibrated temperature, temperature gradient, temperature curvature, and normalized weighted composite, in Valladolid (Spain) for the historic time period from 1867 to 2004, and different AR5 Climate models, when using fourth order polynomials to fit the model results. The highlighted normalized weighted composite is used as the final criteria to rank the climate models according to their capability to describe local climate in Valladolid, with the best model at the top and the worst model at the bottom.

So now we have the missing quality information to interpret the results from different models or combinations between them. Table-2 captures this comparison in terms of the temperature increase we predict for the considered site (Valladolid – Spain) in years 2100 and 2300. as we may see, the results provided by the raw mean of the climate models heavily underestimate the local climate change we can expect in this location, being almost halve of the one we get when using the 'best' model. The results associated to weighting the different models in the basis of their normalized composite residual (quality indicator) reported in Table-1, is also presented in Table-2: It brings about a significant improvement compared with the 'raw' mean of model's results, but why not reporting the 'best' model's results to describe the expected local climate change, mainly when it leads to a significantly higher climate change, if we have identified the relative quality from the different models?, Why to camouflage this 'best' model results under statistical dilution with other 'lower' quality model results?

Perhaps, given the 'quality' from our politicians and their limited capability to take specific action with the available time frames, the conservative approach of diluting the climate model's prediction results under the average (or probability distribution) of non-equal models in terms of quality, could have significant drawbacks.

case	ΔT_{2100} (ºC)	Difference with model's mean in year 2100	ΔT_{2300} (ºC)	Difference with model's mean in year 2300
model's mean	3.7	-	9.5	-
calibrated model's mean	5.2	39.7%	10.9	15.3%
calibrated model's median	5.2	38.8%	11.3	19.2%
calibrated weighted mean (existing)	5.8	54.5%	11.9	26.1%
calibrated weighted mean (incl. best 2100 model)	5.8	54.5%	13.7	44.4%
'best' model (existing)	7.2	93.4%	15.9	67.9%
'best' model	7.2	93.4%	19.1	101.1%

Table-2: Forecast for temperature increase relative to the historic 1867 – 1898 average, in the 2100 and 2300 time horizons, when using different climate model combinations. The values in years 2100 and 2300 are obtained from best fit polynomials capturing the temperature trend evolution. The weighting factors for the corrected mean arise from the compound residuals in temperature, temperature gradient and temperature curvature. The 'best model (existing)' stands for the model that provides the smallest compound residual and that has prediction data for the corresponding year (from the 11 AR5 models, 5 go only till year 2100). The 'best model' adds for year 2300 an estimate of the best overall model (lowest compound residual) which happens to have results just to year 2100, obtained as the ratio of temperature prediction with the most similar model (in terms of results) that goes up till year 2300. The emissions scenario used is the IPCC RCP8.5, with cumulative carbon emissions by year 2100 of 2238 GtC, and CO_2 atmospheric concentration in year 2050 of 540 ppm: This could be considered a BAU scenario, although reality could still unfold worst.

Final note: An expanded version of this analysis will be uploaded at http://xavigarciacasals.blogspot.com.es/