Extended Summary of the Climate Dialogue on regional modelling

Author: Marcel Crok

April 2015

Content

Introduction

Climate models are vital tools for helping us understand long-term changes in the global climate system. Global climate projections for 2050 and 2100 have, amongst other purposes, been used to inform potential mitigation policies, i.e. to get a sense of the challenge we are facing in terms of CO2 emission reductions. The next logical step is to use models for adaptation to climate change as well. Stakeholders have an almost insatiable demand for future regional climate projections. These demands are driven by practical considerations related to e.g. public safety, freshwater resources, especially ecosystems and water related infrastructure, which are vulnerable to climate change.

Hundreds of studies have been published in the literature presenting regional projections of climate change for 2050 and 2100. The output of such model simulations is then used by the climate impacts community to investigate what potential future benefits or threats could be expected. However several recent studies cast doubt whether global model output is realistic on a regional scale, even in hindcast.

So a legitimate question is whether the information from regional climate scenarios for say 2050 or 2100 is reliable enough to be used for all kinds of medium to long term adaptation planning?

We had three excellent participants joining this discussion: Bart van den Hurk of KNMI in The Netherlands who is actively involved in the KNMI scenario's; Jason Evans from the University of Newcastle, Australia, who is coordinator of Coordinated Regional Climate Downscaling Experiment (CORDEX); and Roger Pielke Sr. who through his research articles and his weblog Climate Science is well known for his views on climate model multi-decadal projections. For clarity, both Evans and Van den Hurk are actively involved in regional climate scenarios (decades in the future), Pielke is not.

For personal reasons Evans wasn't able to participate actively in the dialogue after the guest blogs and the first comments were published.

Ready for prime time?

The title of this dialogue was: Are regional models ready for prime time? Another way of asking this is: are regional climate scenarios for say 2050 or 2100 "good" or "reliable" enough to be used for e.g. infrastructural planning decisions? For example, should we increase dikes along our rivers if climate scenarios indicate that extreme rainfall could increase in the coming decades?

The participants had opposed answers to this central question. Pielke's answer to this question was a strong "no". Pielke even said that "by presenting the global, regional, and local climate projections as robust (skillful) to the impacts and policy communities we are misleading them on the actual level of our scientific capability." And also: "I now feel that using the global climate model projections, downscaled or not, to provide regional and local impact assessment on multi-decadal time scales is not an effective use of money and other resources." So for Pielke climate scenarios are misleading policy makers (because they give a false sense of accuracy) and are a waste of research money, i.e. the money could be much better spent.

Van den Hurk's and Evans' answer to the same questions was "yes". Evans for example said: "In the end, climate models are our best tools for understanding how the climate system works. As climate scientists, we will continue to use these tools to improve our understanding of the climate system, and use our understanding of the system to improve these tools. Part of this includes exploring the impact of changing levels of greenhouse gases on the climate by creating future climate projections." And in another comment he wrote: "So RCMs [Regional Climate Models] are not perfect but in many cases are good enough to be useful."

Van den Hurk wrote: "RCMs can be of great help, not necessarily by providing reliable predictions, but also by supporting evidence about the salience of planned measures or policies."

So with regard to the question in the title the views of Pielke are diametrically opposed to those of Evans and Van den Hurk. By reading the whole dialogue and all the comments it becomes clear that this is a case of the glass being half full or half empty. They agree that models still have a lot of problems, also when simulating the past. For Evans and Van den Hurk model projections are nevertheless *useful*, for Pielke they are *useless,* except as model sensitivity experiments and he prefers other approaches to estimate future risks.

Typology

Before we delve deeper into the positions of the participants it is helpful to realise there are different ways in which you can use climate models on a regional scale. In his guest blog Pielke presented four categories. This classification (shown in Table 1) comes from his own papers (Castro et al. 2005, Pielke and Wilby 2012).

Table 1. A Typology of Downscaling Applications

Type	Purpose	Inputs to the Regional Downscaling
1	short-term numerical weather prediction	global analysis of observed data plus observed regional initial conditions
2	regional climate simulation land/atmosphere	information from global or regional reanalyses in which the regional initial conditions are forgotten
3	seasonal prediction	global atmospheric model prediction with prescribed observed surface conditions (e.g., sea surface temperatures)
4	climate prediction	multidecadal global climate model prediction based on initial conditions and prescribed radiative and other climate forcings

Regional climate projections or scenarios, the topic of this climate dialogue, fall in category 4. Pielke explains in his guest blog that the level of skill decreases from type 1 to type 4: “The level of skill achievable deteriorates from Type 1 to Type 2 to Type 3 downscaling, as fewer real world observations are used to constrain the model realism. For Type 4, except for the prescribed forcing (such as added CO2) and the initial conditions, there are no real world constraints.”

Some of the misunderstanding in the dialogue and more general in discussions about the reliability and/or usefulness of regional climate models can be understood from these different types. Pielke supports the use of (regional) climate models to improve our understanding of climate dynamics. He is critical, however, about their value when applied to make type 4 climate predictions, because he has concluded that the models are not yet ready for that task.

Take the following exchange between Evans and Pielke for example. In his guest blog Evans wrote: “A simple rule of thumb is that one can expect downscaling to higher resolution to improve the simulation of regional climate in locations that include coastlines and/or mountain ranges (particularly where the range is too small to be well resolved by the GCM but large enough to be well resolved at the higher resolution) while not making much difference over large homogeneous, relatively flat regions (deserts, oceans,…).”

Somewhat later Evans added: “Above I used simple physical considerations to suggest there would be some added value from regional models compared to global models. Others have investigated this from an observational viewpoint [refs 7,8] as well as through direct evaluation of model results at different scales [refs 9,10,11]. In each case the results agreed with the rule of thumb given earlier. That is, in the areas with strong enough regional climate influences we do see improved simulations of regional climate at higher resolutions.”

So according to Evans downscaling to higher resolution leads to “added value”. Pielke agrees this can be the case, but that the “added value” is limited to type 1-3 applications, where there are real world constraints from observational or reanalysis data. In his guest blog he wrote: “Type 2 downscaling provides an effective way to provide increased value-added information on regional and local spatial scales. It is important to recognize, however, that type 2 downscaling is not a prediction (projection) since the global data comes from a reanalysis (e.g. a reanalysis is a combination of real world observations, folded into a global model).”

So in reaction to the studies Evans mentioned, Pielke said: “These studies are not Type 4 applications when run in a hindcast mode (which is the only scientific way (i.e. the comparison with actual real world observations) to evaluate their skill.”

In his guest blog Van den Hurk also emphasized that you can do more with regional models than just a “prediction”. “It is important to distinguish the various types of analyses that are carried out with RCMs [Regional Climate Models]. And likewise to assess the ability of the RCM to perform the task that is assigned to them. And these types of analyses clearly cover a wider range than plain prediction of the local climate!”

Van den Hurk mentioned a paper by his KNMI colleague Van Haren: “Van Haren et al (2012) also nicely illustrate the dependence of regional skill on lateral boundary conditions: simulations of (historic) precipitation trends for Europe failed to match the observed trends when lateral boundary conditions were provided from an ensemble of CMIP3 global climate model simulations, while a much better correspondence with observations was obtained when reanalyses were used as boundary condition.” In Pielke’s first comment on Van den Hurk’s guest blog he commented on this: “This is an important result. The use of Reanalyses to drive downscaling is a Type 2 application. There is added skill for this type of downscaling. It does not, however, tell us the skill level with Type 4, other than the maximum level of skill possible from Type 4 runs.”

So Pielke emphasizes it should be made very clear what type of application one is talking about. Evans wrote very positively about the value of RCMs: “To date RCMs have been used in many studies12,13 and a wide variety of evaluations of RCM simulations against observations have been performed. […] In most cases the RCMs are found to do a reasonably good job of simulating the climate of the recent past.” Pielke reacted to this saying: “Please provide specific demonstration of the skill of Type 4 downscaling. What has been done, for example, such as by Mearns et al 2012, is to report on Type 2 downscaling results (in which there is added skill) and then to infer this provides the measure for Type 4 downscaling. This is a flawed scientific approach.”

In his guest blog Pielke wrote: “When Type 2 results are presented to the impacts communities as a valid analysis of the skill of Type 4 downscaling, those communities are being misled on the actual robustness of the results in terms of multi-decadal projections. Type 2 results, even from global models used in a prediction mode, still retain real world information in the atmosphere (such as from long wave jet stream patterns), as well as sea surface temperatures, deep soil moisture, and other climate variables that have long term persistence."

Later in the dialogue Van den Hurk did challenge the typology introduced by Pielke. He wrote:

> **How robust is the typology?**
> I agree with Gerbrand [Komen] that Roger's typology is too strict, and that an application dimension should at least be added. Is it designed as a deterministic forecast? Or a probabilistic one? Or one that generates understanding underlying our decisions? Those applications give very different criteria for skill and consistence.

Pielke did not agree saying: "On the typology of downscaling, please be more specific with examples in the literature. I know of no study that cannot be assigned to one of our four types." There were no further challenges from Van den Hurk.

Summary

Pielke is not opposed to the use of climate models per se. They can have added value in the case of the type 1, 2 and 3 downscaling applications. So there is agreement among the participants that regional models can be useful. Pielke warns though for presenting the added value of type 2 downscaling as evidence for the usefulness of doing regional climate predictions 50 or 100 years into the future (type 4).

Skill

The word "skill" was used over 200 times during the dialogue. A returning remark of Pielke was that models first need to show "skill" in hindcast before it makes sense to use them into the future. Pielke claimed models don't have this skill even back in time and his conclusion is that projecting them into the future therefore makes no sense. From Pielke's guest blog:

> Are skillful (value-added) regional and local multi-decadal predictions of changes in climate statistics for use by the water resource, food, energy, human health and ecosystem impact communities available at present?
>
> As summarized in this post, the answer is **NO**.
>
> In fact, the output of these models are routinely being provided to the impact communities and policymakers as robust scientific results, when, in fact, they only provide an illusion of skill. Simply plotting high spatial resolution model results is not, by itself, a skillful product!

During the dialogue it became clear there was no consensus about the definition of "skill". Pielke was most explicit about his definition. *Skill* is defined by him as accurately predicting changes in climate statistics. Or to be more precise: "skill is defined as an ability to produce model results for climate variables that are at least as accurate as achieved from reanalyses. The skill needs to be tested using hindcast runs against i) the average climate over a multi-decadal time period and ii) CHANGES in the average climate over this time period."

Reanalysis data consists of a combination of observations and model output. This is often necessary to check model output because observations alone are not detailed enough to validate the models. Van den Hurk didn't use this definition of skill. He wrote: "I don't know how to assess skill of decadal trends, and so do not require models to reproduce the past trends. A measure of skill of predictions thus should be that the observed climate trends fall within the range of an ensemble of hindcast predictions."

So they used different definitions of skill. The definition of Pielke is much stricter than that of Van den Hurk. Actually when asked about Pielke's definition Van den Hurk said that models are not yet up to that task: "For predictions at the decadal time scale, as Roger identifies in his Type 4 application [i.e. climate scenarios], assessment of skill is actually barely possible. Even a perfect model can deviate significantly from past observed trends or changes, just because the physical system allows variability at decadal time scales; the climate and its trend that we're experiencing is just one of the many climates that we could have had."

So when they would both use Pielke's definition they agree that models have little to no skill. For example Van den Hurk wrote: "I think we should conclude that we agree on the fact that on shorter (decadal) time scales GCM/RCM [Global Climate Models/Regional Climate Models] have shown little regional skill to predict/hindcast observed changes. But that does not necessarily imply that they are useless or have no skill on longer time scales."

In his guest blog Van den Hurk said (our bold):

> For regional climate predictability, the added value of RCMs should come from better resolving the relationship between mean (temperature) trends and key indicators that are supposedly better represented in the high resolution projections utilizing additional local information, such as temperature or precipitation extremes. Also here, **evidence of adding skill is not univocally demonstrated**. Min et al (2013) evaluate the ability of RCMs driven by reanalysis data to reproduce observed trends in European annual maximum temperatures, and conclude that there is a clear tendency to underestimate the observed trends. [...]
>
> Thus indeed, the limitations to predictability or regional climate information by RCMs as discussed by Pielke and Wilby (2012) and others are valid, and care must be taken while interpreting RCM projections as predictive assessments.

With the lack of agreement about the definition of skill the performance of models becomes somewhat subjective. Pielke gave a long list of quotations from papers to show the models have no skill. For example he quoted from Sakaguchi (2012):

> the skill for the regional (5° × 5° – 20° × 20° grid) and decadal (10 – ∼30-year trends) scales is rather limited… The mean bias and ensemble spread relative to the observed variability, which are crucial to the reliability of the ensemble distribution, are not necessarily improved with increasing scales and may impact probabilistic predictions more at longer temporal scales.

Van den Hurk mentions Van Oldenborgh (2013):

> Van Oldenborgh et al (2013) point out that the spatial structure of the mean temperature trend in the recent CMIP5 model ensemble compares fairly well with observations, but anomalies from the mean temperature trend aren't well captured. This uncertainty clearly limits the predictability of temperatures at the regional scale beyond the mean trend.

So Van den Hurk thinks that on a regional scale it should be possible to derive the mean temperature trend. Pielke reacted to this saying: "You need to also report on what level of regional skill of the trends that they achieved. I present an extract from their paper in my guest post." The paper Pielke is referring to here is Van Oldenborgh (2012) instead of Van Oldenborgh (2013). According to Pielke this paper (Van Oldenborgh 2012) reports quite limited predictive skill in two regions of the oceans on the decadal time period, but no regional skill elsewhere. From the paper: "The main source of skill in temperature is the trend, which is primarily forced by greenhouse gases and aerosols. This trend contributes almost everywhere to the skill. Variation in the global mean temperature around the trend do not have any skill beyond the first year."

So although Pielke and Van den Hurk agree that models have very limited skill on the regional scale, even in hindcast, Van den Hurk remains more positive about the performance of models than Pielke. This can be seen in the following exchange (our bold).

> Van den Hurk in his guest blog: "Used this way, also RCMs with **limited predictive skill** can be useful tools for scenario development and providing supporting narratives that generate public awareness or support for preparatory actions."
> Reaction Pielke: "But Type 4 regional downscaling has **NO predictive skill**."

The lack of a mutually agreed definition for "skill" and the related expectations about model performance clearly hampered the discussion.

Summary

The participants did not agree about the definition of skill. Pielke used the strictest definition, saying that model output in hindcast should be at least as accurate as reanalysis data. Van den Hurk used a looser definition, saying that the observed climate trends fall within the range of an ensemble of hindcast predictions. Van den Hurk agreed that using the definition of Pielke, models have no or just very limited skill.
Pielke claims that given the lack of skill in hindcast, it makes no sense to do projections into the future. That such projections are misleading policy makers because they give the illusion of skill. Van den Hurk thinks it's also a matter of time scale. That models have difficulty on time scales of decades because natural climate variability will then dominate. But he believes that on longer time scales the model simulations can still be useful.

Table 2 Definition of skill

	What is your favoured definition of skill?
Pielke	an ability to produce type 4 model results for the statistics (including changes over time) of regional

	climate variables that are at least as accurate as achieved from reanalyses
Van den Hurk	that the observed climate trends fall within the range of an ensemble of hindcast predictions

Predictions, projections and scenarios

There was a lot of discussion about the differences between predictions, projections and scenarios. In his guest blog Van den Hurk wrote: "Scenarios are typically used when deterministic or probabilistic predictions show too little skill to be useful, either because of the complexity of the considered system, or because of the fundamental limitations to its predictability (Berkhout et al, 2013)." A few sentences later he added: "For a scenario to be valuable it does not necessarily need to have predictive skill, although a range of scenarios can be and are being interpreted as a probability range for future conditions."

These remarks are quite crucial. Even though models have limited skill, scenarios are still useful according to Van den Hurk. And a range of model simulations (model ensemble) can be interpreted in terms of probability. Pielke partly agrees with Van den Hurk. He agrees that scenarios can be useful (although not as a prediction but as a model sensitivity assessment); he disagrees that scenarios give a "probability range" for the future.

First the disagreement about "probability". Pielke responded: "They [scenarios] can then be interpreted as a "plausible range" (if they actually are possible). They do not provide a "probability", however, unless one can show skill at predicting probabilities." This leads us back to the discussion about skill. When there is no skill in hindcast, you cannot talk about scenarios in terms of probability, says Pielke. This disagreement about whether scenarios generate a "probable range" was not resolved. In some comments Van den Hurk seemed to agree with Pielke that "probability" is not reachable and the maximum possible is "plausibility". For example when he wrote:

> The purpose of a projection is to depict the possible (plausible) evolution of the system. To my opinion, the process of decision making is not dependent on the (quantitative) predictions provided by climate models, but by the plausibility that the future will bring situations to which the current system is not well adapted. For that a set of imperfect projections is far more useful than a single optimized prediction (which, like a weather forecast with a perfect model, can still be wrong). Decision making does make use but does not entirely rely on this climate information.

Pielke could live with this view as long as one makes clear to stakeholders that the real evolution of the climate can fall outside the range provided by scenarios (our bold):

> I agree with the need to assess what is plausible, but if you agree with the Pielke and Wilby article, than the scenarios that you provide from the downscaled models **may fall outside the range of what actually could occur**. If one insists, they could be included, but there should be a disclaimer given to the policymakers that these regional forecasts have not shown skill when tested in a hindcast mode.

Van den Hurk agreed that there is a chance larger than zero that the real future climate will fall outside the range of model projections but still thinks the models have the order of magnitude of the background climate signal about right (our bold):

> Even with imperfect predictions of regional trends, there is credibility in assuming global warming to be somewhere between 1 and 2 K in 2050. Indeed, there will always be the possibility that – owing to known and unknown model deficiencies and to natural variability – the true warming will be well outside this range, similar to smaller or stronger deviations at the regional scale. But the confidence that this is the right order of magnitude comes from that combination of models, basic understanding of the greenhouse mechanism and involved feedback, and observed trends. Again, we agree that the **probability that the true trends will be outside this range will be larger than zero**, but I do think that the phrase that these climate projections are not credible is not justified.

Pielke disagreed with Van den Hurk's statement about the background climate:

> Where we still disagree, however, is the claim that a long term background climate can be skillfully predicted. What is your basis for this claim? Over the last decade or so, the models have not shown an ability to predict the lack (or very muted) change in the annual average global surface temperature trend. The lower troposphere has similarly not warmed since ~2000 [e.g. see http://www.ssmi.com/msu/msu_time_series.html] When do you assume the warming would restart?

The moderators of the dialogue tried to make the discussion more concrete by talking about the Dutch climate scenarios. In one comment we wrote:

> Let's make it more specific and talk about the KNMI scenario's for The Netherlands. The four scenario's generated so far [Note: these were the KNMI 2006 scenarios] say that The Netherlands will warm 1 or 2 degrees in 2050. Now I would say this is a quantitative prediction/projection. I suppose you would say 1 or 2 degrees of warming is more plausible than 3 degrees of warming or 0 degrees of warming. Is that right? So there is some quantification?

To which Van den Hurk replied (our bold):

> The assumed change of the background climate, as is embedded in for instance the Dutch climate change scenarios (+1 or +2 K increase by 2050), can **be interpreted as a quantitative assessment of how much the climate system will warm**. We derive these values from the spread of IPCC model responses, realizing that these models do have systematic errors, but **reproduce historic variations well enough to have confidence about the approximate range of warming in 2050**. We don't make scenario's for a 0 degree warming, both because these models (and the theory that is embedded in them) show that **this is quite unlikely**, and because making a scenario for a situation that is similar to the present condition does not pose an unforeseen challenge relative to present conditions. But we also do not give scenarios for a 10, or even 5 degrees warming in 2050, because also this is not supported by our understanding of the climate system response, and thus also not by the climate model integrations. Although they are quantitative numbers, these scenarios are not to be

> interpreted as (type 4) forecasts, but as a set of conditions that is relevant to consider when making future assessments.

This is quite confusing because the term "quantitative assessment" suggests there is some "probability" involved, i.e. 1 or 2 degrees of warming is more likely than 0 degrees of warming. Pielke found this claim by Van den Hurk an overstatement, writing:

> Thus, to write that we know "how much the climate system will warm" overstates where we actually are in terms of climate science. This sentence should be written, in my view, as "The assumed change of the background climate, as is embedded in for instance the Dutch climate change scenarios (+1 or +2 K increase by 2050), can be interpreted as ONE assessment of how much the climate system COULD warm."

Van den Hurk then agreed with this rephrasing and saying once again that no probabilities are assigned to the Dutch climate scenarios (our bold):

> With our scenarios we **do not attach probabilities to any of these scenario values**, but we do consider these values to be valuable benchmarks for assessing the impact at regional scale. For that reason, I fully embrace your proposed rephrasing, that the numbers embedded in the Dutch Climate Scenarios can be interpreted as ONE assessment of how the climate COULD warm. But **an assessment that, given the evidence we see, does make sense**.

Still Pielke thinks that even if one assumes a rise in global temperature of 1 or 2 degrees is credible this doesn't mean that regional scenarios are meaningful (our bold):

> Moreover, if we agree the climate models cannot skillfully predict natural variations, why would you conclude that they can skillfully predict changes in climate statistics needed by the impacts and policy communities? The assumption of a global annual average increase in the coming decades +1C and +2C, **is of little use in defining changes in climate impacts at the regional and local scale**, which are so dependent in how large scale circulation features would change in the coming decades. The impacts communities care about such things as the growing season, warm season precipitation, snow coverage, soil moisture, etc, for their locations, not an annual average change in the global surface temperature. These climate metrics are much more dependent on the large scale circulation features than a global annual average surface temperature trend.
>
> My bottom line is that while the global climate models, when run with added CO2 and other greenhouse gases, show that this is a warming effect, they are inadequate tools to assess the consequences of these human climate forcings on the regional and local scale. Even worse, the models inadequately include the diverse myriad effects of aerosols and land use/land cover change on the climate system, so they are already hindered in their ability to accurately represent the real world spectrum of human climate forcings.

Earlier on Pielke did say that regional climate scenarios could be useful. Pielke thinks they are useful in the sense of what he calls a "process" or "sensitivity study". A process study is (our bold):

> Process studies: The application of models to improve our understanding of how the system works is a valuable application of these tools. The term "sensitivity study" can be used to characterize a process study. In a sensitivity study, a subset of the forcings and/or feedback of the system may be perturbed to examine its response. The model of the system might be incomplete and not include each of the important feedbacks and forcings. Marshall et al. (2004) is one example of this type of model application. This is **also the type of modeling being used to make climate projections (predictions) in climate assessments such as the IPCC reports** (e.g., see Pielke 2002a). This application of the models is made despite their inability to show multi-decadal regional and mesoscale skill in forecasting changes in climate statistics when run in a hindcast mode (e.g., see Pielke 2013, and also Section 13.5). These IPCC model runs are actually process model simulations.

So for Pielke climate projections or scenarios, as done by the IPCC and by national climate institutes, are "sensitivity studies" that can improve our understanding of how the system works. As he confirmed in a comment that was directed towards the German climate scientist Hans von Storch:

> It is fine, of course, to use a different word (i.e. "projection" instead of a "sensitivity study") as long as the meaning is clear and they mean the same thing. The IPCC, unfortunately, has failed to properly communicate that the "projections" are only providing, at best, what is possible, and does not in any way, provide a demonstration of quantitative likelihood or evidence of any actual regional and local prediction skill on multi-decadal time periods.

Projections and predictions

This brings us also to the difference between "predictions" and "projections". In his guest blog Pielke said these are just synonyms:

> One issue that we need to make sure is clearly understood are the terms "prediction" and "projection". The term "projection", of course, is just another word for a "prediction" when specified forcings are prescribed; e.g. such as different CO2 emission scenarios - see Pielke (2002). Thus "projection" and "prediction" are synonyms.

However Evans and Van den Hurk disagreed. Evans wrote in his first comment (our bold):

> Roger presents a case for abandoning climate model projections altogether – or at least not allowing the impact/adaptation community and policy makers to see them as they will gain "an erroneous impression on their value." In my experience **this is certainly the case if you talk about the simulations as predictions rather than projections** – the climate models are not predicting what the weather will be on the 5th of May 2051 – they are providing projections of the climate based on emission scenarios and initial conditions. When dealing with those outside the climate science community I have certainly found the distinction to be important.

Van den Hurk wrote in his first comment:

> Roger considers projections and predictions to be synonyms, but I disagree with this, and also feel that this assumption is an important element in the controversy under discussion. A prediction is a statement on the expected situation in the near future starting from a current state of the system, while a projection is a possible evolution of the system given an assumed (external) driver.

Pielke explained why he thinks predictions and projections are the same:

> First, the attempt to discriminate between a "prediction" and a "projection" is not appropriate. The use of the term "projection" with respect to added greenhouse gases, for example, is just a "what if" prediction (i.e. given that forcing). These models can be (and must be) tested in hindcast runs to assess their level of skill at predicting what actually occurred. This skill must be shown before the results of future runs on multi-decadal time scales (as "projections" can be trusted).

In a later comment Pielke mentioned the definitions for climate prediction and projection of the IPCC (from AR4):

> Climate prediction – A climate prediction or climate forecast is the result of an attempt to produce an estimate of the actual evolution of the climate in the future, for example, at seasonal, interannual or long-term time scales. Since the future evolution of the climate system may be highly sensitive to initial conditions, such predictions are usually probabilistic in nature.
>
> Climate projection – A projection of the response of the climate system to emission or concentration scenarios of greenhouse gases and aerosols, or radiative forcing scenarios, often based upon simulations by climate models. Climate projections are distinguished from climate predictions in order to emphasize that climate projections depend upon the emission/concentration/radiative forcing scenario used, which are based on assumptions concerning, for example, future socioeconomic and technological developments that may or may not be realised and are therefore subject to substantial uncertainty.

Pielke then commented: "Clearly a "climate projection" IS a "climate prediction, using the IPCC definitions, when the emission/concentration/radiative forcing scenario are known. This can be done using hindcast model runs, and these model runs can be quantitatively tested. The source of much of this confusion appears to be in the assumption by the IPCC and others, that while weather prediction is an initial value problem, climate prediction (on multi-decadal time periods) is a boundary value problem; the term "projection" then being reserved for the later."

Gerbrand Komen, the former research director of KNMI and familiar with climate scenarios, in a comment tended to agree with Pielke: "I tend to agree with Roger that there is no sharp distinction between predictions and projections. They are both modelled constructions of the state of the climate system at a future time."

There was no further discussion on this distinction, so it cannot be said that the issue was resolved.

Komen also introduced the term "belief" in the discussion in the following way (our bold):

> The real question is about the **belief** in the possibility that these predictions may verify. This is not black or white, because the prediction may be approximately correct for some variables and wrong for some other variables, and because one can distinguish different degrees of beliefs about the likelihood that a prediction will (or might) come true [very uncertain, possible, very unlikely, . . very likely . .]. Roger seems to assume that skill in hindcasts is necessary for usefulness. I can see that skill strengthens our **belief** in the likelihood, but in my essay [http://bit.ly/cCb1n2 or http://home.kpn.nl/g.j.komen/Uncertainties.pdf] on Rogers website I have made two points that may be relevant for the present discussion: 1. One should realize that there is ALWAYS a chance that predictions do not come true, even if the model has shown skill in hindcast studies; 2.There are a number of tests (but more than just the skill in a hindcast) that feed our belief in the usefulness of a model, indeed in a sort of Bayesian manner.

Pielke didn't like the term "belief" as he regards it as unscientific:

> If the predictions are inaccurate in hindcast studies, they certainly should not be claimed to be able to have skill in any projections for the coming decades. Of course, a prediction, even if it shows skill in a hindcast set of runs, can still be inaccurate in the future. But a NECESSARY condition to accept a model prediction as accurate is that they show skill.
>
> On your second comment, what tests are you referring to? Also, the use of the word 'belief' is not a robust scientific demonstration of added value. Please be more specific.

Komen explained the term belief doesn't have a religious meaning: "The word 'belief' as used here should not be seen as 'religious belief', but rather as a technical term used by philosophers, psychologists, and social scientists. There is an extensive body of literature from such fields as the philosophy of science and a field called social epistemology, that study the role of belief in science. It is my guess that this is also relevant for the present discussion."

But for Pielke the issue is quite black and white: "My Reply – I suggest we defer using the word "believe". The model results should be accepted as robust ONLY when model predictions (i.e. the hypotheses) cannot be shown to be falsified with real world data. This is the scientific method. I have presented peer reviewed papers that do, in fact, falsify the models even with respect to their ability to predict (in hindcast) the current climate."

Van den Hurk is closer to Komen's view:

> The use of the word "belief" in the discussion tends to give it another twist: to what sense can we "believe" that altering climate statistics might give rise to reconsider some of our society decisions when we have nothing more than (imperfect) models to feed this "belief"? Andries Rosema even finds it "frightening" that generating pictures of how climate statistics could change is feeding these considerations, as they introduce a large risk of alarmistic messages and overinvestments in mitigation or adaptation measures. I tend to turn the argument around: ignoring the possibility that the climate could look different in the future

> while making these society decisions, is given the evidence we get from a large source of scientific studies (including, indeed, imperfect models) a bit naive.

Van den Hurk then referred again to a paper he had written with Prof Frans Berkhout "that analyses the extent to which one can use plausible (model-derived) storylines for decision making processes in society." In this paper they "discerned a number of mind frames that form the starting point for this decision making process". These frames are derived from a book by Rumelhart (1989):

> * Reasoning by similarity: a problem is solved by seeing the current situation as similar to a previous one in which the solution is known. In this category fall intuition, reasoning by example or experience, generalization and analogical reasoning.
>
> * Reasoning by simulation: a problem is solved by imagining the consequence of an action and making explicit the knowledge that is implicit in our ability to imagine an event. This category includes story-building to mentally simulate the events leading up to a certain ending.
>
> * Formal reasoning: a formal symbolic system, such as logic or mathematics, is employed in the solution of a problem. Examples are formal mathematical models of biophysical or social systems, including climate scenarios.

Van den Hurk thinks the disagreement between Pielke and himself is the result of Pielke following "formal reasoning" while he follows "reasoning by simulation":

> Roger (and Andries) tend to stress the need for formal reasoning, while I (and Gerbrand) make a plea for the second way of reasoning. Roger could reply again by stating that models that don't show skill in projecting changing statistics cannot be used for this reasoning by simulation, but I remain to disagree with him: the skill of climate models to project changing climate statistics at decadal time scales can formally not be established due to large role of natural variability, but is also not always required for generating useful information that enters the imagination process.

Pielke found this distinction too vague: "However, you did not define what is to be the output of such scenarios. I suggest that what the impacts and policymakers need to start with is an identification of what are the consequences of specified changes in climate statistics."

Summary

"For a scenario to be valuable it does not necessarily need to have predictive skill, although a range of scenarios can be and are being interpreted as a probability range for future conditions." This sentence is nice summary of the many issues that were dealt with in this section.

Pielke disagreed mostly with this statement. As long as models have no skill in hindcast one cannot attach probabilities to projections or scenarios with these models. Pielke therefore emphasized that the real climate could fall outside the range given by the projections or scenarios and that policy makers should be aware of that. Van den Hurk confirmed that no probabilities are attached to the different Dutch climate scenarios.

For Pielke the distinctions between predictions, projections and scenarios are inappropriate. Scenarios and projections are just "what if" predictions and should be tested in hindcast. Evans and Van den Hurk disagreed. For them a projection is a possible evolution of the system given an assumed (external) driver. It differs from a prediction which is a statement on the expected situation in the near future starting from the current state of the system.

Pielke and Van den Hurk seemed to agree that scenarios can be interpreted as a "plausible range" and not as a "probable range". Probability is not necessary for Van den Hurk to be useful: "To my opinion, the process of decision making is not dependent on the (quantitative) predictions provided by climate models, but by the plausibility that the future will bring situations to which the current system is not well adapted."

Top down versus bottom-up

A key issue in the dialogue was the distinction between a "top down" and "bottom-up" approach. Pielke criticizes the IPCC for focussing solely on the top down approach. From his guest blog (our bold):

> This [the bottom-up approach] is a more robust way of assessing risks, including from climate, than using the approach adopted by the Intergovernmental Panel on Climate Change **(IPCC) which is primarily based on downscaling from multi-decadal global climate model projections**. A vulnerability assessment using the bottom-up, resource-based framework is a more inclusive approach for policy makers to adopt effective mitigation and adaptation methodologies to deal with the complexity of the spectrum of social and environmental extreme events that will occur in the coming decades as the range of threats are assessed, beyond just the focus on CO2 and a few other greenhouse gases as emphasized in the IPCC assessments.

The IPCC starts with projections with global climate models (most recently CMIP5), which are then downscaled to the regional level (for many regions in the world). The output of these regional projections are then used in impact models. The bottom-up approach reverses this process:

> As a more robust approach, we [Pielke and Wilby] favor a bottom-up, resource-based vulnerability approach to assess the climate and other environmental and societal threats to critical assets. This framework considers the coping conditions and critical thresholds of natural and human environments beyond which external pressures (including climate change) cause harm to water resources, food, energy, human health, and ecosystem function. Such an approach could assist policy makers in developing more holistic mitigation and adaptation strategies that deal with the complex spectrum of social and environmental drivers over coming decades, beyond carbon dioxide and a few other greenhouse gases.

Van den Hurk is also in favour of this bottom-up or vulnerability approach but he thinks that doing regional climate scenarios still has a function in this approach (our bold):

> I fully embrace Pielke's plea for a system analysis that takes the vulnerability of the system as a starting point. But from this kind of analyses, **frequently the stakeholders are the**

> **participants that ask for support from (regional) climate models to illustrate the possible alternative future conditions**.

Pielke thinks there are cheaper and less time consuming instruments to use within the bottom-up approach. One can for example use historical, paleo-record and worst case sequences of climate events. Pielke: "Added to this list can be perturbation scenarios that start with regional reanalysis (e.g. such as by arbitrarily adding a 1C increase in minimum temperature in the winter, a 10 day increase in the growing season, a doubling of major hurricane landfalls on the Florida coast, etc). There is no need to run the multi-decadal global and regional climate projections to achieve these realistic (plausible) scenarios."

Elsewhere he said that when one is looking for plausible states of the future climate, doing regional climate scenarios is unnecessarily costly (our bold):

> Running the models to produce "plausible" scenarios is **reasonable**, but it is **very expensive** in terms of time and other resources. Even more importantly, unless they can actually be shown to be "plausible", it is not appropriate to present to the impacts community without **the disclaimer** that they have not shown skill at predicting the climate metrics of interest when the models are run in hindcast.

So they agree that the bottom up vulnerability approach is very useful for policy makers. Pielke believes that in addition one could do perturbation experiments with impact models to paint plausible futures. That is much cheaper than running climate scenarios. Van den Hurk believes regional climate scenarios do help to project plausible future climate states and are therefore worth the time and money.

> But at longer time scales the change of the probability of a very rare extreme event is dominated by the change in the background climate, where small changes in the mean can lead to large changes in the probability of extremes (see e.g. the scaling between surface temperature and 99 percentile precipitation by Lenderink et al (2011) or the non-linear relation between trends and extremes by Rahmstorf and Coumou (2011)). Again, superposition of this changing background climate on projections with high resolution models resolving the relevant process does give valuable insights of the possible changes in risks and vulnerability.

Summary

Pielke criticizes the "top down" approach used by the IPCC, in which global models are downscaled to the regional level and then used as input for impact models. He is in favour of a "bottom-up" approach in which one starts with the vulnerability of important societal resources like water, food, energy, human health and ecosystems for all kind of pressures (including from climate).

Van den Hurk agrees with this approach but thinks that regional climate scenarios can still play a useful role within this bottom-up approach. Pielke thinks there are cheaper options available, like the use of historical data. In addition one could do perturbation experiments in which one tests for example what would be the effect of a 1 or 2 degrees of warming on these important resources.

References

Berkhout, F., B. van den Hurk, J. Bessembinder, J. de Boer, B. Bregman andM. van Drunen, Framing climate uncertainty: using socio-economic and climate scenarios in assessing climate vulnerability and adaptation, Regional and Environmental Change, 2014, 14, 879-893.

Castro, C.L., R.A. Pielke Sr., and G. Leoncini, 2005: Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS). J. Geophys. Res. - Atmospheres, 110, No. D5, D05108, doi:10.1029/2004JD004721. http://pielkeclimatesci.wordpress.com/files/2009/10/r-276.pdf

Lenderink, G., H.Y. Mok, T.C. Lee and G.J. van Oldenborgh, Scaling and trends of hourly precipitation extremes in two different climate zones – Hong Kong and the Netherlands Hydrology and Earth System Sciences, 2011, 4701-4719, doi:10.5194/hessd-8-4701-2011.

Mearns, Linda O. , Ray Arritt, Sébastien Biner, Melissa S. Bukovsky, Seth McGinnis, Stephan Sain, Daniel Caya, James Correia, Jr., Dave Flory, William Gutowski, Eugene S. Takle, Richard Jones, Ruby Leung, Wilfran Moufouma-Okia, Larry McDaniel, Ana M. B. Nunes, Yun Qian, John Roads, Lisa Sloan, Mark Snyder, 2012: The North American Regional Climate Change Assessment Program: Overview of Phase I Results. Bull. Amer.Met Soc. September issue. pp 1337-1362.

Pielke Sr., R.A., and R.L. Wilby, 2012: Regional climate downscaling – what's the point? Eos Forum, 93, No. 5, 52-53, doi:10.1029/2012EO050008. http://pielkeclimatesci.files.wordpress.com/2012/02/r-361.pdf

Rahmstorf S and Coumou D 2011 Increase of extreme events in a warming world Proc. Natl Acad. Sci. USA 108 17905–9

Rumelhart DE (1989) Toward a microstructural account of human reasoning. In Vosniadou S, Ortony A (Eds.) Similarity and analogical reasoning. New York, NY, Cambridge University Press, pp. 298-312

Sakaguchi, K., X. Zeng, and M. A. Brunke (2012), The hindcast skill of the CMIP ensembles for the surface air temperature trend, J. Geophys. Res., 117, D16113, doi:10.1029/2012JD017765.

Van Oldenborgh, G.J., F.J. Doblas-Reyes, B. Wouters, W. Hazeleger (2012): Decadal prediction skill in a multi-model ensemble. Clim.Dyn. doi:10.1007/s00382-012-1313-4

Van Oldenborgh GJ, Doblas Reyes FJ, Drijfhout SS, Hawkins E (2013) Reliability of regional climate model trends. Environmental Research Letters 8