TODAY'S TOPICS IN SCIENCES

Climate prediction: an evidence-based perspective

Richard Goody · Guido Visconti

Abstract This paper considers climate prediction from the perspective of the experimental, physical sciences, and discusses three ways in which the two differ. First, the construction of long-term climate series requires benchmark measurements, i.e., measurements calibrated in situ against international standards. An instrument capable of accurate, benchmark measurements of thermal, spectral radiances from space is available but has yet to be used. Second, objective criteria are needed to evaluate measurements for the purpose of improving climate predictions. Techniques based on Bayesian inference are now available. Third is the question of how to use suitable data to improve a climate prediction, when they are available. A method based on the Bayesian Evidence Function is, in principle, available, but has yet to be exploited. None of these three aspects are considered in current operational climate forecasting. All three are potentially capable of improving forecasts, and all are subjects of current research programs, with the likelihood of their eventual adoption.

Keywords Climate predictions · Bayes inference · Climate benchmarks · Climate modeling

R. Goody
School of Engineering and Applied Science,
Harvard University, Cambridge, MA 02138, USA

G. Visconti (✉)
Department of Physics, Center of Excellence
on Integration of Remote Sensing Techniques and Numerical
Modeling for the Forecast of Severe Weather (CETEMPS),
University of L'Aquila, 67010 L'Aquila, Coppito, Italy
e-mail: guido.visconti@aquila.infn.it

1 Introduction

The physical sciences fall into two broad categories: *experimental sciences* (physics, chemistry) which rely on the simplest class of evidence, the controlled experiment; and observational sciences (earth, atmospheric, ocean sciences and astronomy), evidence for which consists of observations of vast, complex, and uncontrollable systems. Both take an approach to the natural world that depends upon evidence to eliminate unsatisfactory hypotheses.[1] Nevertheless, the approach of an individual investigator differs greatly whether confronted with a controlled experiment or a huge, uncontrollable system.

Climate forecasting is a part of the atmospheric and oceanic sciences but with an added complication due to the use of large digital computers for environmental modeling.[2] An environmental model attempts to couple together all relevant physical and chemical components of a natural system in the context of a fluid flow model. The model is subject to changing external conditions (*forcings*) and projects to future states in many small time steps. Environmental models can be of daunting complexity, and they can give predictions of great detail. According to Beven (2009), they engender a point of view previously unknown to science, to which he gives the name Pragmatic Realism. Pragmatic Realism sees the output of an environmental model as a form of reality that can be brought closer to an ultimate reality by improving the model components. It rests on a disputable view of reality (see Beven 2009, Chapter 2) and is a further departure from the experimental model of physical science.

---

[1] See Gauch (2003) for a thoughtful discussion of uses of evidence in science.

[2] See Beven (2009) for an account of environmental modeling.

108

Rend. Fis. Acc. Lincei (2013) 24:107–112

Climate predictions are performed by environmental models known as Global Climate Models or GCMs, and they involve large uncertainties. These uncertainties have been fastened upon by critics of global climate change, and action on climate change has significantly slowed in recent years, partly for this reason. It is by no means clear that the political opposition to climate change would disappear if the uncertainties in climate prediction were substantially reduced. Nevertheless, that should be the aspiration of the climate community and, for the most part, it is.

In recent decades, the experimental sciences have had extraordinary successes, from controlling nuclear energy to reading the genetic code, and their findings command great respect in virtually all quarters. Because of their simpler systems, experimental scientists can limit uncertainties in their evidence. They aim for measurement accuracy that can distinguish between alternate hypotheses, and they vary parameters to produce a body of evidence. The purpose of this paper is to ask whether there may be lessons from experimental science that might help improve the quality of climate predictions. Uncertainties in climate predictions will not be eliminated in this way, but we do believe that it could be a step in the right direction. The common link is the best use of evidence, and the difference in this respect is that for the experimental scientist evidence dominates all other considerations, but for climate predictions the emphasis is on Pragmatic Realism.

2 Climate predictions and their uncertainties

The subject of climate change and climate prediction is now defined by the encyclopedic publications of the U.N. Intergovernmental Panel on Climate Change. The latest edition was issued in 2007 (IPCC AR4 2007) and a new edition is expected soon. The relevant volume is that by Working Group I. This volume has 152 lead authors from 30 countries with 650 correspondents and represents substantial ranges of opinions of almost all of the climate community.

For this paper, the most important results of IPCC AR4 are the 100-year predictions of global surface temperature (Fig. 1). There are 24 independent predictions by different national groups, all using the same, plausible forcing by increasing concentrations of greenhouse gases. The spread of the IPCC climate temperature trends is 2.2–3.5 K per century, a range sufficiently large to influence the climate change debate. This spread is largely caused by differences between models. Another source of uncertainty is the year-to-year fluctuations on the records in Fig. 1. This is caused by non-linear, fluid instabilities in the models (*natural variability*). Natural variability also occurs in the real
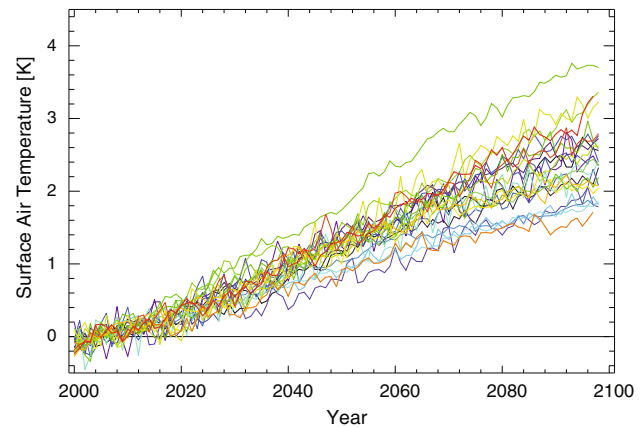

Fig. 1 The IPCC AR4, ensemble of climate predictions. Climate predictions of globally averaged surface temperature by 24 independent GCMs as reported in IPCC AR4, with some additions. The climate trends (*slopes*) cover the range 2.2–3.5 K

climate in the form of weather which, on time scales up to several years, can have a very large amplitude.

There are additional uncertainties, for example, the forcing is controlled by unpredictable political and economic factors. Other uncertainties are related to the assumptions made by the different models. And we must assume that the mechanisms of climate will be the same in 100 years time as they are now, which we cannot know with certainty. We have no data on the future climate that might allow us to identify and modify sources of error, and these projections have the status of hypotheses, rather than confirmed scientific conclusions.

Part of the uncertainty in a projection is due to uncertainty in the environmental model, the GCM. Unlike the quality of the projection, the quality of the model can be objectively assessed by performing predictions from the past to the present, and comparing the result to present data. Within the limits imposed by natural variability and the accuracy of the data, it is possible, in principle, to develop models which account completely for all evidence available at the time the prediction is made. Other things being equal we may reasonably expect that the better the model, the better the prediction. This is the point at which a greater emphasis on data could prove to be valuable.

Chapter 8 of IPCC AR4 (2007), *Climate models and their evaluation*, describes many efforts to evaluate climate models. For the most part, these investigations are for special purposes and not directed towards the systematic evaluation of model quality. This question is, however, contained within a discussion of metrics. On page 591, we read "*The possibility that metrics based on observations might be used to constrain model projections of climate change has been explored for the first time, through the analysis of ensembles of model simulations. Nevertheless a proven set of metrics that might be used to narrow the*

range of plausible climate projections has yet to be developed". This is the part of current climate research that deserves more emphasis. It is expected that there will be many competing ensembles of climate predictions in the future. A metric could enable the best to be selected.

We shall examine some of the issues involved in developing a useful metric of model performance, and the best data for this purpose. All atmospheric variables are potentially useful, and there are also combinations of variables which might be more valuable than the individual variables alone. Consequently, there are a very large number of observations available for use in a metric, most of which have not yet been used for this purpose. We notice that model intercomparison projects (AMIP, CMIP) have found large difference between models and data.

3 The quality of climate data

As shown in Fig. 1, the predicted change of surface temperature is approximately 3 k in a century. In order to make discriminating tests between theories, measurements should, preferably, be made with an accuracy of about 0.1 K, and this accuracy must be sustainable over centuries. Weather data are not of this quality, even when *re-analyzed*, a process by which all available data types are made internally consistent in the context of a weather prediction model. The only way to achieve such accuracy is by means of regular instrument calibrations, while in use, against international standards, *benchmark measurements*. It is very difficult to make benchmark measurements with instruments carried on meteorological radiosondes, but it is possible to do so in the more benign conditions that exist in space. Space observations have the additional advantage that the entire surface of the earth can be measured with the same instrument, which is evidently important for a global phenomenon.

There are currently three satellite benchmark measurements. One measures the total solar radiation in terms of electrical standards using an active cavity radiometer (Willson and Helizon 1999). A second measures air density in terms of international standards of time, by means of occultations of the radio signals from Global Positioning System satellites (Kursinski et al. 1997). We shall briefly describe the third instrument, which has not yet flown, a radiometer calibrated in terms of international temperature standards (Anderson et al. 2004).

Figure 2 shows the spectrum of radiation leaving the Earth to space. This is important to climate studies for two reasons. First, the total outgoing radiation is the response of the planet to the incident solar radiation. This is one of the most fundamental relationships governing climate change. Second, the complex features shown in Fig. 2 are caused by
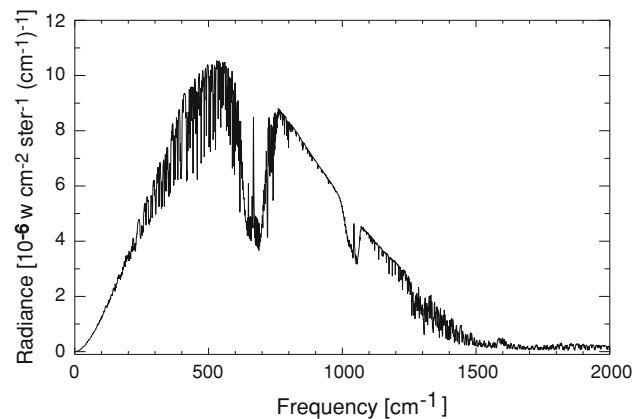
Fig. 2 Spectrum of the heat radiation leaving planet earth. The many small features are the signatures of greenhouse gases

Fig. 3 Schematic of the CLARREO interferometer. There are two independent instruments boresighted on the same region of the surface. There are *four independent black bodies*; a minimum of one would suffice

molecular absorptions, and they can be unraveled to yield information on surface and air temperatures, clouds, and gaseous concentrations although with coarse resolution.

The spectrum in Fig. 2 is recorded by an interferometer, the purpose of which is to compare the Earth's radiation to that from a calibrated cavity (a *black body*). A schematic for the CLARREO instrument is shown in Fig. 3; this instrument is a miniature standards laboratory. One important aspect of this instrument is redundancy. There are two independent instruments performing the same function, and each instrument has two independent black
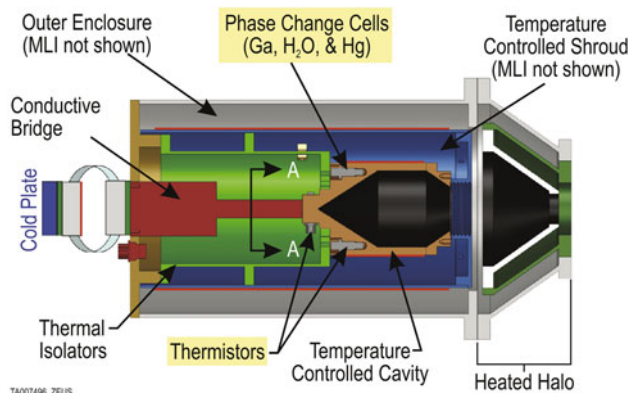
110

Rend. Fis. Acc. Lincei (2013) 24:107–112

Fig. 4 A CLARREO blackbody. The phase change cells of gallium, water and mercury ensure the benchmark character of the spectrometer

Fig. 5 Bayesian inference: a schematic illustration

Fig. 6 Evolution of a prediction. Model predictions of surface temperature trends over 50 years are modified by data at 10, 20, 30, 40, and 50 years. Calculations for two data types are shown: surface temperature trends (tas), and trends in all satellite radiances taken together (radiances) (Huang et al. 2011)

bodies, for a total of four, where only one is required. The purpose of this redundancy and other features of the instrument design is to ensure that the link between Earth and black body remains the same, and the onus for a benchmark result then rests squarely on the accuracy of the black body (Fig. 4).

Good design of a black body can ensure that its accuracy is solely determined by the accuracy with which the cavity temperature can be determined. This temperature is measured in terms of the liquid to solid phase changes of water, gallium, and mercury. The international standard scale of temperature is defined in terms of phase changes, and the CLARREO thermometers are, therefore, consistent with that scale. The thermometers might fail, but they cannot give an incorrect reading.

4 Selecting and using new evidence

Not all climate measurements are equally useful for improving climate predictions, and the ability to choose between data types on the basis of value is useful both for allocating observing resources and for the efficient use of computer resources.

But before discussing this topic, we need to take note of the fact that both predictions and observations are uncertain to some degree. When we compare them we are comparing two uncertain quantities and our conclusions will also be uncertain. The appropriate mathematical framework for handling such quantities is the science of statistics, a discipline with its own assumptions and methods, and there is increasing belief that, for assimilating climate data, the appropriate system of statistical inference is that of Bayes (Sivia 1996; Beven 2009). Bayesian inference treats relationships between probability density functions (pdfs), quantities which express the probability of occurrence of a
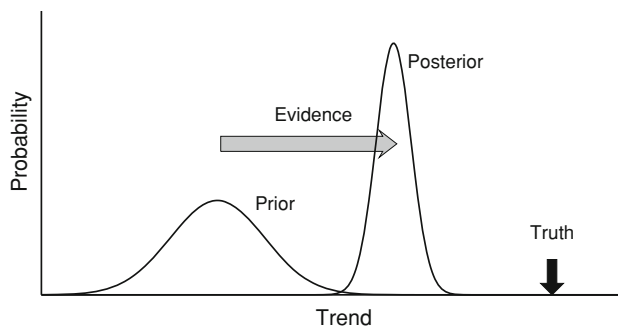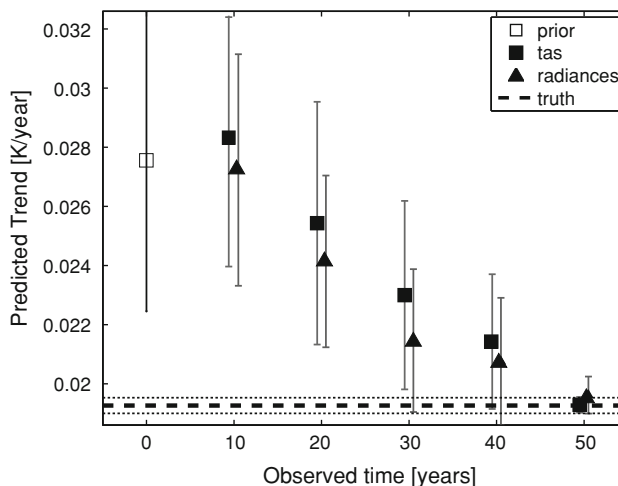
quantity in terms of the value of that quantity (see Fig. 5 for examples of pdfs of climate trends). With only 24 results in Fig. 1, it is only possible to describe the simplest possible pdf (a gaussian pdf) which has the same mean and the same standard deviation (the spread of the pdf) as the data. The most likely prediction and its uncertainty are correctly represented by a gaussian pdf.

The pdf of the data in Fig. 1 is known as the prior, i.e., the information that exists prior to the introduction of new evidence. Now suppose that new evidence becomes available (the horizontal arrow in Fig. 5), this adds to our knowledge about the climate, and requires that the prior be modified to the posterior. The maximum of the posterior must be closer to the truth than the prior, because it is based on more evidence, and it should be more certain (smaller spread). Bayesian inference allows these statements to be placed on a quantitative basis, see Huang et al. (2011) and Sexton et al. (2012).

Figure 6 shows the results of a 50-year prediction of globally averaged surface temperatures after adding information about the climate of the first 10 years, the first

The Evidence Function

$$P(d \mid M) = \int_{-\infty}^{\infty} P(d \mid x)\, P(x \mid M)\, dx$$

x = climate forecast; d = data

M_1 is a more credible model than M_2 if

$$\frac{\text{Evidence Function }(M_1)}{\text{Evidence Function }(M_2)} > 1$$

Fig. 7 The Evidence Function $P(\mathbf{d}|M)$ is the probability that data **d** would be obtained given the model M. Note that **d** is a vector of data types representing anything from all available data types to a single data type

20 years, etc. The maxima of the pdfs are shown by the points and arrowheads and the standard deviations by the vertical lines. The mean of the prior is represented by the empty square point; the filled square and triangle points and lines are the posteriors. Two types of data are added: squares are for globally averaged surface temperatures; triangles are for certain radiances, measured from a satellite. The figure demonstrates the expected improvement of accuracy and reduction of uncertainty as new evidence is added, and shows, surprisingly, that adding space radiances may be slightly more useful than adding surface temperature data, even when the latter is the predicted quantity. These calculations can be extended to include many data types. Data types that stand out as being more useful than others for improving predictions of surface temperatures are to be preferred.

When suitable data have been obtained, they must be used to develop a metric suitable for routine selection between models. Huang et al. (2011) point out that such a metric falls out of the Bayesian analysis outlined in the previous section. The Bayesian metric is the Evidence Function, $P(\mathbf{d}|M)$, or the probability that the evidence, **d**, is compatible with a model, M. Figure 7 shows an expression for the Evidence Function. It is presented here as a comparison between two models. It is hard to assign a meaning to the probability of a single model unless there is something to measure it against. **d** is written as a vector and can represent all available data or any selection from it. Different selections may be appropriate to different users of a climate prediction (recreational, business, military, etc.), because not all users have the same interests, and each model may be better in some respects than others.

It will not be straightforward or easy to develop the Evidence Function as a metric for the quality of an ensemble of climate models. It will require arduous numerical research to accomplish this goal. However, Bayesian inference establishes that such a metric exists and that it is probably the best metric available.

5 Conclusions

In the previous sections, we have shown how some data ideal for climate research are becoming available, and how data in general can be used to select the best available climate ensemble. These prior ensembles can then be upgraded to improved posteriors, in the manner demonstrated in Fig. 6. We have not discussed the use of data for improving models rather than their predictions. Model improvement involves improving individual process models that are coupled together in a GCM. Parts of the climate system are isolated and methods used that are familiar to an experimental scientist (see Garratt 1992, for a monograph on the atmospheric boundary layer, and Goody (1964) on atmospheric radiation; both topics are treated as process models in GCMs). The use of data to study the GCM predictions is different, and involves a significant gap between an experimentalist and the ideas of Pragmatic Realism. Our discussion shows how this gap can be partially closed, and the Bayesian approach that we employ may be the best that can be done.

Research described in IPCC AR4 (2007) shows that there are climate scientists who are aware of these issues and are working on them, but these are not the most active areas of climate research. The reason for this can probably be traced to the importance of ideas of Pragmatic Realism, and to the fact that climate research has prospered under its aegis and has risen to political prominence[3] without considering additional complications. Moreover, the experimental view of science which we offer is evidently not the only road to important advances in scientific knowledge.[4] Perhaps it should be asked whether experimental method is in any way appropriate to the climate problem?

Given these considerations, how and why should the proposals that we outline be followed? First the "why". Bayesian inference indicates that an increased use of better evidence can only increase the quality of climate predictions; whether by much or by little has to be determined. This path needs to be pursued, partly for the sake of scientific integrity and partly because it could increase confidence in climate predictions, a focus for concern at the present time. The "how" is more difficult, because it may involve a trade-off with established programs. Fortunately, it appears that thinking within the climate community may already be taking new directions. Informal sources indicate that funding has been made available for research into

---

[3] The work of the IPCC was recognized with the award of the 2007 Nobel Peace prize.

[4] The theory of Evolution is known as one of the most important contributions of science to human knowledge, but was developed without an experimental basis.

Bayesian metrics, and that more than one research group is working on the problem.

Climate observing systems are complex and extremely expensive. It is unlikely that a climate network based on benchmark measurements will ever be funded. The most important source of operational climate data is the established international meteorological networks, which consist of both orbiting and in situ measurements. When natural variability is large, these networks are probably sufficiently accurate for climate research; but if the importance of natural variability is decreased through the availability of long time series, this may not continue to be the case. Re-analysis has been seen as a solution to inadequacy of the weather networks, but there has recently been a realization that re-analysis may not deal with small, slowly developing errors, of the kind that can only be eliminated with benchmark measurements. Benchmark measurements might be used, not as a climate network in themselves, but to calibrate the re-analyses of meteorological data. This would be an appropriate compromise between the search for more reliable climate predictions and budget realities.

If these advances in climate techniques take place, the lessons on offer from experimental science will have been assimilated.

References

Anderson JG, Dykema JA, Goody RM, Hu H, Kirk-Davidoff DB (2004) Absolute, spectrally-resolved, thermal radiance: a benchmark for climate monitoring from space. J Quant Spectrosc Rad Trans 85:367–383

Beven K (2009) Environmental Modeling: an Uncertain Future, Routledge, London, p 309

Garratt JR (1992) The atmospheric boundary layer. Cambridge University Press, Cambridge, p 316

Gauch HD (2003) Scientific method in practice. Cambridge University Press, Cambridge, p 435

Goody RM (1964) Atmospheric radiation: I theoretical basis. Oxford University Press, Oxford, p 436

Huang Y, Leroy S, Goody RM (2011) Discriminating between climate observations in terms of their ability to improve an ensemble of climate projections. PNAS 108:10405–10409

IPCC AR4 (2007) Climate change 2007—the physical science basis. Cambridge University Press, Cambridge, p 996

Kursinski ER, Hajj GA, Schofield JT, Lindfield RP, Hardy KR (1997) Observing earth's atmosphere with radio occultation measurements using the global positioning system. J Geophys Res 102:23429–23465

Sexton DMH, Murphy JM, Collins M, Webb MJ (2012) Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. Clim Dyn 38:2513–2542

Sivia DS (1996) Data analysis; a Bayesian tutorial. Oxford University Press, Oxford, p 189

Willson RC, Helizon RS (1999) EOS/ACRIM III instrumentation. Proc SPIE 3750:233–242