A methodological note on the making of causal statements in the debate on anthropogenic global warming

Jarl K. Kampen

Received: 9 February 2010 / Accepted: 4 October 2010 / Published online: 19 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract At best, the empirical evidence for human impact on climate change, more specifically, the anthropogenic global warming (AGW), is based on correlational research. That is, no experiment has been carried out that confirms or falsifies the causal hypothesis put forward by the International Panel on Climate Change (IPCC) that anthropogenic increasing of green house gas concentrations very likely causes increasing of the (mean) global temperature. In this article, we point out the major weaknesses of correlational research in assessing causal hypotheses. We further point out that the AGW hypothesis is in need of potential falsifiers in the Popperian (neopositivistic) sense. Some directions for future research on the formulation of such falsifiers in causal research are discussed. Of course, failure to find falsifying evidence in empirical climate data will render the AGW hypothesis much stronger.

1 Introduction

“Most of the observed increase in global average temperatures since the mid-twentieth century is very likely due to the observed increase in anthropogenic green house gas concentrations, and it is likely that there has been significant anthropogenic warming over the past 50 years averaged over each continent except Antarctica” (IPCC 2007: p. 5). In this proposition, which was part of the summary for policy makers of the famous IPCC report on climate change which has influenced politicians throughout

the world, the phrase “very likely due to” needs clarification. As such, it is a causal statement of the form “increase of X very likely causes increase of Y.” More specifically, the causal statement is “anthropogenic increasing of GHG concentrations very likely causes increasing of the (mean) global temperature.”

According to Bunge (1959; see Tacq 2010), a causal relationship between X and Y requires that (a) the relationship is conditional (if X, then Y); (b) unique (one cause, one effect); (c) asymmetrical (when X causes Y, then Y does not cause X); and (d) invariable (no probabilistic causality). Although elegant in its simplicity, application of Bunge's rigid framework does not seem to expand very much beyond the realm of elastic collisions as observed, e.g., in a game of billiard. On the other end of this rigid definition stands American pragmatism, where one ceases to look for metaphysical grounds of a theory but instead, measures the value of a hypothesis on causality based on the cash-value of the idea (James 1907):

Grant an idea or belief to be true, what concrete difference will its being true make in anyone's actual life? How will the truth be realized? What experiences will be different from those which one would obtain if the belief were false? What, in short, is the truth's cash-value in experiential terms?

Since as a matter of fact, the cash-value of the idea of AGW has proved to be positive, this alone may serve as sufficient motivation for its maintenance in science. For those satisfied with this argument, *Q.E.D.*, and the story ends here. However, others may not dismiss the fact so easily that causal statements in science have been subject to debate for as long as science exists (see, e.g., Pearl 2000; Tacq 2010). As early as in 1737, the philosopher David Hume wrote that

J. K. Kampen (✉)
Environmental Policy Group/Research Methodology Group,
Wageningen University and Research,
Hollandseweg 1,
6706 KN Wageningen, The Netherlands
e-mail: jarl.kampen@wur.nl

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connection; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other.

In other words, regardless of the rationale on which a causal statement has been based, its validity will always remain to be open for dispute. While early scientists depended on the principle of verification for the validation of theories (including causal hypotheses), the critique of Hume gradually convinced scientists that other principles should be applied in order to establish whether or not a theory was scientific. Whereas the positivists of the Wiener Kreis still advocated that scientific theories are sound if they can be verified by empirical data, the logical shortcomings of the verification criterion led Karl Popper to radically brake with this tradition.

Briefly stated, the major shortcoming of the verification criterion is that it allows only experience to decide upon the truth or falsity of scientific statements (Popper 1965: 42; see Rapp 1975). Popper's most important contribution to the debate was to state that every scientific theory should be able to list counter-examples which, if found in reality, disconfirm (“falsify”) the theory. This is the principle of falsification. In the case of anthropogenic global warming (AGW), the theory should list one or more counter-examples that could (potentially) disconfirm the theory. This listing of potential falsifiers appears to be missing in the present debate on AGW. In fact, some skeptics in the debate on AGW point out that all natural climatic disasters are used as evidence (verification) for the human impact on climate, whereas evidence that a post WWII global warming is absent in, e.g., the Greenland Ice-Core Bore Record is ignored as falsifying evidence (see, e.g., Dahl-Jensen et al. 1998; Feldman and Marks 2009). Needless to say that a methodologically sound theory would encompass all available evidence and not “cherry-pick” those pieces of evidence that confirm the theory while ignoring those that do not.

Unfortunately, when a theoretical phenomenon such as AGW becomes a global political program, it soon becomes vulnerable to methodological fallacies in the realm of social and political science. Leaving aside the quality of used data and methods, the IPCC report aimed at reaching a consensus. Consensus is recognized by some social scientific methodologists as the defining feature of social science (Swanborn 1996; Feyerabend 1987). However, if reaching consensus were really the hallmark of sound science, the scientific theories of Galileo, Copernicus, Darwin, and many others would never have seen daylight. Also, there is no guarantee that majorities will reach sensible opinions (think only of the democratic Weimar republic in the 1930s). Finally, scientists

need to make a living, and they will not bite the hand that feeds them, an argument used by some advocates of AGW who claim that climate skeptics are sponsored by “Big Carbon”. Therefore, consensus must be dismissed as a defining feature of science. The IPCC recognizes the limitation of consensus by adding the phrase ‘and much evidence’ when it makes statements as in, e.g., “there is high agreement *and much evidence* that with current climate change mitigation policies and related sustainable development practices, global GHG emissions will continue to grow over the next few decades” (IPCC 2007: p. 7, italics added). We must therefore discuss the sources of evidence that are used to formulate the many causal statements on AGW issued in the report.

The quality of all scientific research depends of course, on the quality of the data that are being processed. Regardless of the quality of the (statistical) model used for analysis, if bad data are fed to the model, then the result of the analysis will be bad. This principle is known as garbage in–garbage out. In other words, if the data that are fed into climate models are open to dispute, then so are the projections of these models. In the scientific (i.e., peer-reviewed) literature, several authors have expressed doubts about the quality of the analyzed data and the possibility to derive at valid inferences on human impact on global warming (e.g., Jaworowski 1994; Soon et al. 2004; Michaels 2008; Pielke et al. 2007). However, since the author of this article is no expert on climate science, the issue of whether or not data used in climate science are of enough quality will be left for others to decide. Instead, in this methodological note on the making of causal statements in the debate on AGW, we focus on the study designs that are used to establish the causal hypotheses. The following sections discuss briefly the consequence of a lack of experiment and the relying on correlational data for establishing causal relationships. This discussion prepares the ground for the formulation of possible falsifiers of AWG. Some concluding remarks remain in the last section.

2 On the establishing of causality in time series

2.1 The consequence of the lack of experiment

The challenge in corroborating any causal hypothesis is to determine what kinds of evidence constitute actual *proof* of the hypothesis. The study design appropriate for establishing causality, that is for establishing cause and effect relationships is the experiment, as any freshman's course or introductory work in research methodology will reveal (e.g., Kumar 2005: 100; Ford 2000: 141; de Vaus 2001: 70; Gomm 2008: 60; Neale and Robert 1986: 134). The key

to the conducting of true experiments is the installing of control groups that receive no, or different, experimental interventions. However, despite efforts to find planets with equal conditions in the universe, no replicate of the Planet Earth has been found at the time that this article went to press and therefore, no true experiment can be conducted that would compare climate change on a planet with and without anthropogenic carbon dioxide production. Furthermore, artificial interventions where for instance, the impact of doubling anthropogenic carbon dioxide on global warming is investigated are if not infeasible, surely unlikely to be carried out by contemporary researchers. These observations may seem rather straightforward, but their consequences are far from trivial, for the immediate implication is that the only study design that is available to test the AGW hypothesis is the longitudinal study, either a panel or a trend study (see, e.g., Kumar 2005: 111).

Lack of a control group in experiments make the testing of counterfactual relationships, for instance $\neg X \rightarrow \neg Y$ ("if there were no Industrial Revolution, then there would not have been an increase in global temperature"), impossible.

In sciences where experimentation is difficult if not impossible, such as the social sciences, researchers rely on correlational research in order to establish causal relationships. Simply stated, two variables correlate positively when increasing values of X correspond to increasing values of Y . Thus, a change of X leads to a change of Y , which is one of the prerequisites for establishing causality (see Section 1). In order to find a (statistically significant) correlation, however, there is no requirement that X precedes Y . Both hypotheses $X \rightarrow Y$ and $Y \rightarrow X$ will produce identical correlations.

Furthermore, a significant correlation of empirical data can be observed when, in fact, no causal relationship is involved. For instance, the priests of ancient Egypt were paid for seeing to it that, in the evening, the sun would set and in the morning, the sun would come up again. Because they upheld their ceremonies every evening and morning, and because every morning the sun did come up and every evening the sun did set, the priests proved to the general public that their actions had value. Of course, the causal relationship between the making of offerings and the behavior of the sun was nil, whereas the correlation between the actions of the priests and the sun equaled one, that is, the correlation was maximal.

Another well-known problem relates to the so-called third variable (e.g., Neale and Robert 1986: 239). In order to illustrate this phenomenon, a much used example is where the number of churches in a city is compared with the number of crimes committed in a city: with a large enough sample of cities, a positive correlation will be found indicating that churches cause crime (or vice versa). The third or *intervening* variable, in this case, is the size of the

city, and the empirical correlation is called *spurious*. In sum, even when correlation is a necessary condition in causality, it is not a sufficient one.

2.2 Granger causality

In the literature, the notion of Granger causality has received increasing interest (e.g., Dufour and Renault 1998), an econometric concept that has also been applied in the analysis of climate data (e.g., Stern and Kaufmann 1999; Verdes 2005; Kauffmann et al. 2006). Briefly stated, $X \rightarrow_g Y$ that is a variable X is said to Granger cause another variable Y if the observation of X up to time t can help one to predict Y at time $t+1$, when the corresponding observation on Y is available. That is, G-causation implies that knowledge of the past of X and Y produces better predictions of Y than knowledge of the past of Y alone (Freeman 1983). This definition of causality does not require presence of a plausible theory for the causal connection. Rather, it focuses on one important aspect of causal hypotheses, namely the power to predict, and the validity of a predictive model can be expressed in terms of error in prediction and model fit statistics (e.g., likelihood ratios).

Use of Granger causality in order to confirm the AGW hypothesis has received critique in the literature (e.g., Triacca 2001, 2005). The most important problem with statistical models such as Error Correction Models and related time series models is that they cannot serve as proof of a causal relationship. Firstly, because the associations that are found may be spurious, meaning that one or more intervening variables have been omitted in the model. Of course, such intervening variables can be included in the model, e.g., in a vector Z of auxiliary variables (which requires of course, measurement of this vector), but proving that all intervening variables have been captured is impossible. This is because absence of proof is not equivalent to proof of absence.

Secondly, how does one prove that X is the precursor for Y in a statistical analysis of empirical data? At the very least, the goodness-of-fit statistics of the models where both data generating processes ($X \rightarrow_g Y$ and $Y \rightarrow_g X$) are assessed need to be compared. This comparison can have four possible outcomes with corresponding conceptual and operational hypotheses:

Conceptual hypothesis:	Operational hypothesis:
1. $X \rightarrow Y \wedge \neg (Y \rightarrow X)$	$X \rightarrow_g Y$ has significant model fit and $Y \rightarrow_g X$ does not
2. $\neg (X \rightarrow Y) \wedge Y \rightarrow X$	$X \rightarrow_g Y$ does not have significant model fit and $Y \rightarrow_g X$ does
3. $\neg (X \rightarrow Y) \wedge \neg (Y \rightarrow X)$	Neither $X \rightarrow_g Y$ nor $Y \rightarrow_g X$ fit the data
4. $X \rightarrow Y \wedge Y \rightarrow X$	Both $X \rightarrow_g Y$ and $Y \rightarrow_g X$ display significant model fit

In applied research, the operational hypothesis relates to the expected outcome from statistical inference based on analysis of empirical data. It is important to note that validity of the operational hypothesis provides evidence for, not proof of, validity of the conceptual hypothesis. For instance, the first outcome provides evidence for the simple mechanistic hypothesis that X is a cause (or at least a mediator) of Y ; and similarly, the second outcome provides evidence for the alternative hypothesis that Y causes (or is a mediator of) X . Outcome 3 is a falsifier for both outcome 1 and outcome 2. The fourth outcome suggests the possibility of a feedback loop, i.e., causal recursion.

Modeling matters in terms of verification (and falsification) become complicated in the case that a feedback loop is present. Among issues that need to be taken into account are the different time lags for the $X \rightarrow Y$ and $Y \rightarrow X$ relationships; the time span that is covered by data on X and Y ; the direction of the feedback (negative or positive); and the (non)linearity of the relationships. In verifying causal recursion, it is further important to note that outcome 4 is a necessary but insufficient condition for presence of a feedback loop. For instance, consider two simple time series $X = \sin(t)$ and $Y = \cos(t)$. In these data, a feedback loop is absent, while both $X \rightarrow_g Y$ and $Y \rightarrow_g X$ will produce equally well fitting models: we need only remember that $\cos(t) = \sin(t + \pi/2)$ and $\sin(t) = \cos(t - \pi/2)$. Also, in Error Correction Models such as the one applied by Kauffmann et al. (2006: p. 256), if the model produces significant parameters with Y as effect (i.e., dependent variable), it will at the same time produce significant parameters with X as effect (this is so because the mathematical model specifies the two variables in terms of linear combinations of each other). In both cases, the data will result in outcome 4, and other kinds of evidence than the verification of statistical model fit are needed to decide what the proper direction of the causal relationship is (see, e.g., Hausman 1982). A possible strategy for obtaining such evidence is outlined in the next section.

2.3 Qualitative model validation and falsification

By verifying causal relationships, we only see one side of the coin when, in fact, we need to see both sides in order to show that the coin is genuine. That is to say, a verifying modeling procedure must be supplemented by a falsifying one. It was suggested by Van der Zouwen and Van Dijkum (2001: 237) that the formulation of potential falsifiers in recursive models can be based on qualitative mathematical knowledge:

A qualitative strategy may result in the identification of nullclines and equilibria which put constraints on

the set of (differential) equations and make it possible that whole sets of models are falsified and others verified.

Qualitative mathematical knowledge is equally useful in the validation of relatively simple constituent non-recursive models, that is, causal chains that are parts of the fully specified model.

For example, we may establish that X indeed precedes Y (and not vice versa) by placing constraints on the pattern of minima and maxima of the time series (see also Van Dijkum et al. 2001). Formally, we may require that first (and higher order) derivatives of the cause predict, with a given time lag, first (and higher order) derivatives of the effect:

$$\frac{\partial}{\partial t} X \rightarrow_g \frac{\partial}{\partial t} Y.$$

Notably, if, in a given time span, maxima (minima) of X are not consistently followed (let alone preceded) by maxima (minima) of Y , then the hypothesis that X is a precursor for Y may be considered falsified. This, in turn, requires a time span of measurements where changes of the gradients must be detected for both X and Y . Furthermore, the time series of derivatives must be computed independently for X and Y (artifacts are produced if they are derived from differentiation of the autoregressive models used for establishing G-causality!), for instance, by the use of splines (see, e.g., de Boor 1978).

Obviously, the qualitative strategy of model validation requires a theory about causal relationships that can be expressed in a mathematical model. It also requires data of substantial quality. For instance, if the AGW hypothesis is to be confirmed, data are required that quantify the amount of human produced CO_2 as separated from the amount produced by other sources (e.g., volcano's) and as separated from the amount of CO_2 dissipated in oceans for a time span that covers theoretically justified time lags (perchance in the order of centuries). An “aggregate” causal variable that combines radiative forcing of greenhouse gasses, anthropogenic sulfur emissions, and radiative forcing of solar irradiance cannot provide “direct evidence that, since 1870, human activity is largely responsible for the increase in global surface temperature” (cfr. Kauffmann et al. 2006: 225, 250). Of course, evidence that the simple non-recursive causal chain is valid (conceptual hypothesis) mounts with increasing numbers of maxima and minima that follow the regularity constraint (operational hypothesis). However, the precise point when evidence changes into proof remains to be a subjective decision based on consensus among scientists.

3 Concluding remarks

The fact that a true experiment (one including control groups) is not feasible in the case of studying causality relating to warming on Planet Earth does not mean that the weaknesses of correlational research do not apply in climate research. A well-known weakness of such research designs refers to fact that absence of proof (of spurious relationships) does not imply proof of absence (of spurious relationships). In order to scientifically corroborate the AGW hypothesis, the present focus on verification of the AGW hypothesis should shift towards a focus on its falsification. A potential falsifier is when empirical data fail to show that maxima (minima) of the cause (e.g., human produced CO₂) produce, at a specified time lag, maxima (minima) of the effect variable (e.g., global temperature). The latter requirement necessitates the availability of time spans of data that are large enough to display changes in the gradients of both cause and effect variables, and the application of models that allow for inference on derivatives.

As said, the challenge in corroborating any causal hypothesis is to determine what kinds of evidence constitute actual *proof* of the hypothesis. The field of verification and falsification of recursive spatial-temporal causality is underdeveloped and merits future research. Meanwhile, consensus among scientists will remain to play a great role in deciding when empirical evidence suffices as proof of a causal hypothesis. However, in adopting Swanborn's (1996) "regulative idea of striving after truth by consensus within the scientific community over research results", we are always in danger of replacing the purpose of science (knowledge) with a by-product of science (consensus in the form of "common sense"). Karl Popper's requirement of a sound scientific theory, that it should produce counter-examples that falsify its validity, serves as a first line of defense against this danger. Of course, failure to find falsifying evidence in empirical data will render the AGW hypothesis much stronger.

Acknowledgements This article has benefited enormously from comments and suggestions given by Jacques Tacq (HUBrussel, Brussels, Belgium), Cor van Dijkum (Utrecht University, The Netherlands), and an anonymous reviewer of TAC. Needless to say, any remaining errors are the sole responsibility of the author.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Bunge M (1959) Causality. The place of the causal principle in modern science. Harvard University Press, Cambridge, MA
- Dahl-Jensen D, Mosegaard K, Gundestrup N, Clow GD, Johnsen SJ, Hansen AW, Balling N (1998) Past temperatures directly from the Greenland ice sheet. *Science* 282:268–279
- de Boor C (1978) A practical guide to splines. Springer Verlag, Berlin
- de Vaus D (2001) Research design in social research. Sage, London
- Dufour J, Renault E (1998) Short run and long run causality in time series theory. *Econometrica* 66(5):1099–1125
- Feldman S, Marks V (2009) Global warming and other bollocks. John Blake Publishing, London
- Feyerabend P (1987) Farewell to reason. Biddles Ltd., London
- Ford ED (2000) Scientific method for ecological research. Cambridge University Press, Cambridge, UK
- Freeman JR (1983) Granger causality and the time series analysis of political relationships. *Am J Polit Sci* 27(2):327–358
- Gomm R (2008) Social research methodology. A critical introduction. Palgrave-Macmillan, New York
- Hausman D (1982) Are there causal relationships among dependent variables? *Philos Sci* 50:58–81
- IPCC (2007) Summary for policy makers. Retrieved from the internet (08.02.2010) from www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-spm.pdf
- James W ((1907 reprinted 1981)) Pragmatism: a new name for some old ways of thinking. Hackett Publishing, Indianapolis, IN
- Jaworowski Z (1994) Ancient atmosphere-validity of ice records. *Environ Sci Pollut Res* 1(3):161–171
- Kauffmann RK, Kauppi H, Stock JH (2006) Emissions, concentrations, and temperature: a time series analysis. *Clim Change* 77:249–278
- Kumar R (2005) Research methodology. Sage, London
- Michaels PJ (2008) Global warming: correcting the data. *Regulation* 31(3):46–52
- Neale JM, Robert RM (1986) Science and behavior. An introduction to methods of research. Prentice-Hall, New Jersey
- Pearl J (2000) Causality. Models, reasoning, and inference. Cambridge University Press, New York
- Pielke R, Nielsen-Gammon J, Davey C, Angel J, Bliss O, Doesken N, Cai M, Fall S, Niyogi D, Gallo K, Hale R, Li H, Raman S (2007) Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull Am Meteorol Soc* 88(6):913–928
- Popper K (1965) The logic of scientific discovery. Routledge, London
- Rapp H (1975) The methodological symmetry between verification and falsification. *Z Allgemeine Wissenschaftstheorie* 6(1):134–139
- Soon WH, Legates DR, Baliunas SL (2004) Estimation and representation of long-term (>40 year) trends of Northern-Hemisphere-gridded surface temperature: a note of caution. *Geophys Res Lett* 31(3):1–4
- Stern DI, Kaufmann RK (1999) Econometric analysis of global climate change. *Environ Model Softw* 14:597–605
- Swanborn P (1996) A common base for quality control criteria in quantitative and qualitative research. *Qual Quant* 30(1):19–35
- Tacq J (2010) Casuality in qualitative and quantitative research. *Qual Quant*. Available at Online First: www.springerlink.com/content/n330814063203313/fulltext.pdf
- Triacca U (2001) On the use of Granger causality to investigate the human influence on climate. *Theor Appl Climatol* 69(1):137–138
- Triacca U (2005) Is Granger causality analysis appropriate to investigate the relationship between atmospheric concentration of carbon dioxide and global surface air temperature? *Theor Appl Climatol* 81(3–4):133–135
- van Dijkum C, Lam N, Ganzeboom HBG (2001) The dynamics of educational expansion: a simulation model. In: Geyer RF, van der Zouwen J (eds) Sociocybernetics: complexity, autopoiesis, and observation of social systems. Greenwood Publishing Group, Westport
- van der Zouwen J, van Dijkum C (2001) Towards a methodology for the empirical testing of complex social cybernetic models. In: Geyer RF, van der Zouwen J (eds) Sociocybernetics: complexity, autopoiesis, and observation of social systems. Greenwood Publishing Group, Westport
- Verdes PF (2005) Assessing causality from multivariate time series. *Phys Rev* 72:1–9