From Uncertain Time-Series Data Can One Get A Deterministic Extrapolated Trend?

Franco Pavese

(formerly Research Director in Metrology at the National Research Council,
then at Istituto Nazionale di Ricerca in Metrologia (INRiM), Torino 10135, Italy)

E-mail: frpavese@gmail.com. Profile at: https://researchgate.net/profile/Franco-Pavese/

From partial knowledge to inference and back

There is an extremely vast literature on this subject matter with very different solutions for this difficult problem, based on mathematical/statistical model, neural network, hybrid models, …

It is a basic rule in Science to make a distinction between *theoretical studies*, mind constructions based on inference, and *experimental studies*, based generally on the acquisition of numerical values by means of experimental setups—mostly human-made. The latter can be a preliminary step, or can follow the former one, but in both cases it acts as a *validation* of the former, within a given *uncertainty*.

Uncertainty is therefore an *unavoidable* component that makes the gained knowledge *partial* (exceptions exist, e.g., after Centuries, today about the roundness of the Earth!).

The theoretical studies are generally expressed in the symbolic language of mathematics. They are called "laws" when assumed to describe a general feature of Nature, or they are called "models" when their function is merely to track experimental findings. Both *laws and models* are assumed to be consistent with the experimental findings (considered an "evidence") within the uncertainty of the latter.

However, often *no uncertainty is associated to them*, except when they are "data-driven", i.e. they are based only on datasets (**B. Knusel, Ch. Baumberger, Understanding climate phenomena with data-driven models, in Studies in History and Philosophy of Science 84 (2020) 46-56**).

Most often today a model is *not* data-driven in the above sense, but arises instead from an *inference* based on the underlying "laws" governing the observed phenomena.

Here comes the issue of "diversity" in science, arising again from the fact that knowledge is generally partial.

However, there is a basic difference between a *diversity of experimental results* in different datasets and a *diversity of inferences* in theoretical (mathematical) models.

In the case of data, though they are individually considered as "external facts" (objective) contributing to extend or confirm our knowledge, diversity may arise from their level of variability in a dataset or from inconsistencies between datasets, both due to "errors" or to experimental setup imperfections or to limited capability. ***These factors are taken as the origin of data uncertainty***. Though most of these influence factors might originate from human mind, however **uncertain data are still considered as objective facts**.

In the case of theoretical inference, diversity arises instead from different "positions" of scientists in evaluating which of possible underlying laws may be included in the model, and their respective relationships and cross-influence. Thus, the construction of such a model have to be considered fully a *mind construction* (inference), related to human perception of the external world, only via its consistency with (part of) the available experimental (uncertain) data.

Here comes another feature attributed to science, especially from frames outside science: society, politics, …: **the ability to provide a forecast** about the future validity of the assumptions underlying the observed trends of the present models, on the lack of change in the influence factors. Concerning the laws, according to Popper's principle of falsification, an *a posteriori* evaluation tool, a forecast of a *law must* assume *ipso facto* its future validity, differently from the case of a *model*. However, the basic questions are:

1) **How from uncertain time-series data can one get a deterministic extrapolated function**?
2) **Can an inference contribute to an increase of the current partial knowledge**?

Only climate-related examples will be used in the following, due to their importance today.
See, e.g.:
Christoph Baumberger, Reto Knutti, Gertrude Hirsch Hadorn, Building confidence in climate model projections: an analysis of inferences from fit, WIREs Clim. Change 2017, 8re454;
IPCC, Evaluation of Climate Models. WG1ARS, Chapter 09;
Mastrandrea, M.D., C.B. Field, T.F. Stocker, O.Edenhofer, K.L. Ebi, D.J. Frame, H. Held, E. Kriegler, K.J. Mach, P.R. Matschoss, G.W. Yohe, and F.W. Zwiers: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. 2010 IPCC

Inference uncertainty

This issue is vastly treated in the literature, especially in applications to *economic forecast*, where decisions have to be taken based on the inference of future trends.
Limiting us here to *experimental science* and leaving apart the case of data-driven method (see later), the inference is made by means of a model consisting of mathematical function(s), i.e. using a *deterministic tool.* It is chosen:

(a) to "best fit" the data representing the current knowledge, or
(b) to use a (set of) functions (called "model") inferred to represent (be the cause of) the observed (underlying) physical trend, by best reproducing the current data trend.

In both cases, an uncertainty *must* be associated to inference, arising from (i) the data uncertainty, (ii) the degree of confidence in the correctness/suitability of the inference.
The two cases require very different tools for its evaluation, being the first basically of random nature, while the second derives from assumed systematic effects.

(a) Uncertainty of data best fit and confidence in forecast

The suitability and credibility of the function used for the fit is most generally evaluated by means of the standard deviation of the fit, care being also taken to avoid overfitting.
However, the uniqueness of the fit is not ensured, thus the differences between fitting functions can also contribute to fit uncertainty: this is more critical when the function is to be extrapolated ahead, where no constraints exist because of the lack of data. The current partial knowledge is instead extrapolated ahead by usually assuming that the fitting-function trend is still represented by the current function trend, or via suppressing-old/introducing-new functions from inference of modified future influence factors.

(b) Uncertainty from model parameters, and forecast credibility

Except in the case of data-driven inference (see below), forecast can only be performed by using continuous functions, since data are obviously not (yet) available.
What is the meaning of the uncertainty of a function? It is not uniqually defined.

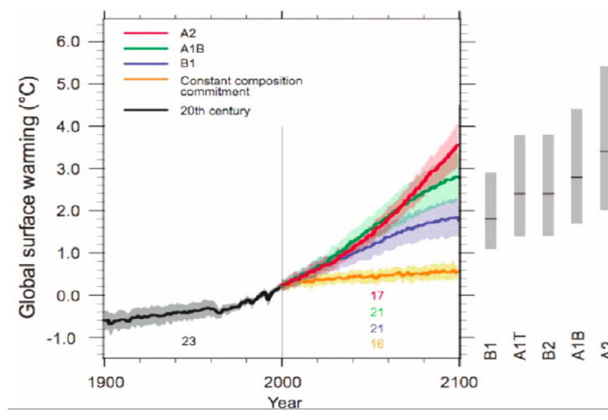
(i) Function parameters are assumed having a sort of "flexibility" (sensitivity analysis);
(ii) Data ahead are simulated several times and the function changed behaviour is recorded;
(iii) The dispersion of the trends of functions arising from different inferences is analised.

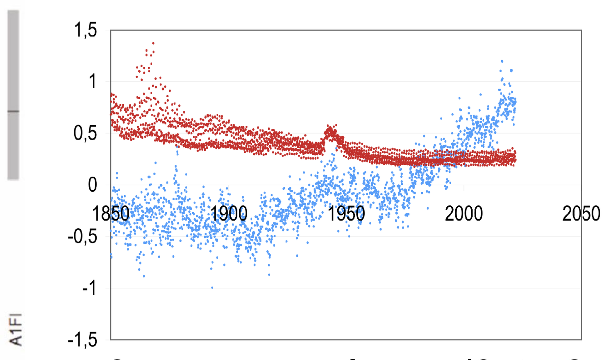
(i) This is called the "sensitivity method", of the function trend to a variation of each parameter across a given range of values. It allows to show the different level of criticality of the parameters (or of a single function-component of the inference).
(ii) New sets of data are inferred ahead and the suitability of the function is then evaluated, both with probabilistic tools e.g., non-parametric ones, like bootstrap, MonteCarlo, …).
(iii) Several functions are inferred, in two ways: by using the same inferred influence factors, when this non-uniqueness is possible; or, by inferring (partially) different influence factors.
This method is popular and used, for example by IPCC in climate analyses ("In cases where observations are lacking, we resort to inter-comparison of model results to provide at least some quantification of model uncertainty *via inter-model spread*"). Strictly speaking this method does *not* concern the uncertainty of any specific function.

In this respect, a *deterministic* function, as a process-based model normally approximates the data only according to its own mathematical representational capability—except the completely different case when the function is aimed instead at accurately tracking the data trend, e.g. when the aim is to exactly track a (geometrical) "profile".
So normally no deterministic model can ensure an *exact* description of an experimental data set.

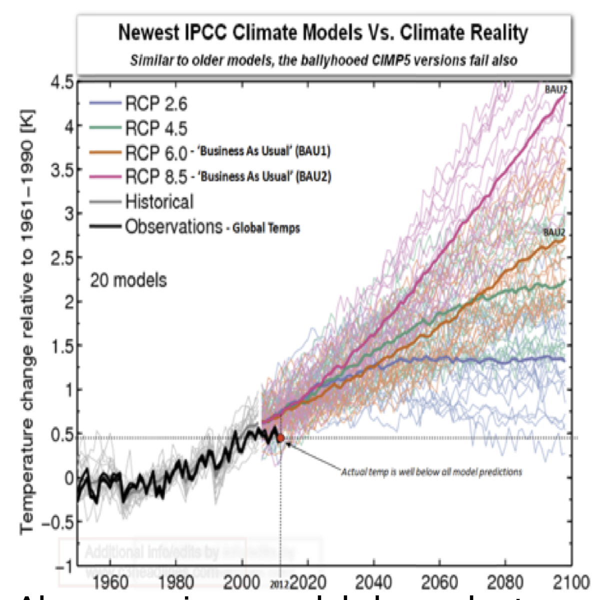
(c) Uncertainty of data-driven inference

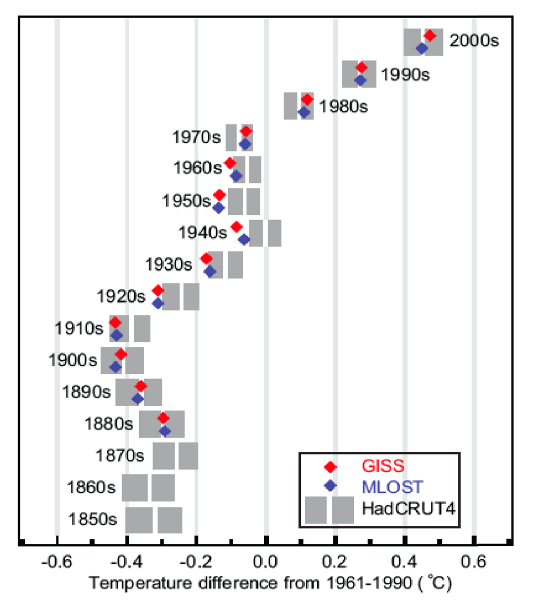
This kind of inference makes use of non-parametric statistical methods and is not expressed by functional relationships (see reference above) like in the previous case (called "process-based" models). The method is specifically used by means of "machine-learning" algorithms, "driving" the best adaptation of subsequent sets of inferred data to specific rules obtained through a "training" process. It is a machine-driven method of type (b) above.
The paper from B. Knuesel and C. Baumberger cited above is a good start point, with its references, to understand the details of this technique, where the authors conclude: "we agree that both model interpretability and the lack of evidence linking the model to the target pose difficulties for the fitness of data-driven models for understanding. However, we argued here that neither problem necessarily precludes data-driven models from serving as vehicles for understanding in specific instances. Creating data-driven models in situations where sufficient background knowledge is available to argue from the coherence of the model with background knowledge to its representational accuracy can provide exactly the kind of evidence that reduces the link uncertainty discussed by Sullivan (2019). Advances in explainable machine learning can generally be expected to further increase the fitness of data-driven models for the purpose of understanding as they will increase the graspability of data-driven models".
However these authors miss at all to tackle the issue of the data uncertainty.





SAMT estimates from HadCRUTS



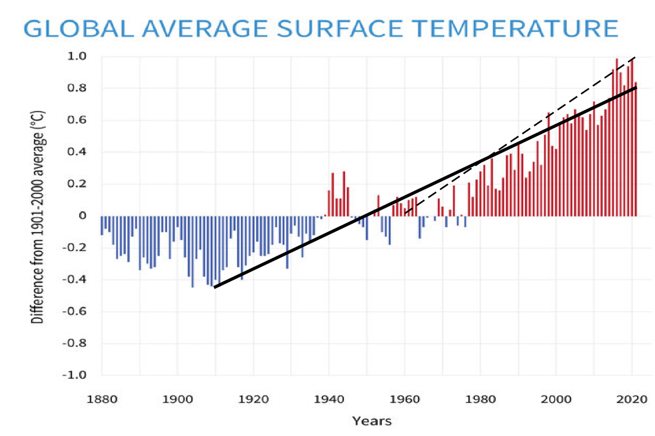


Above: various model-dependent SAMT estimates

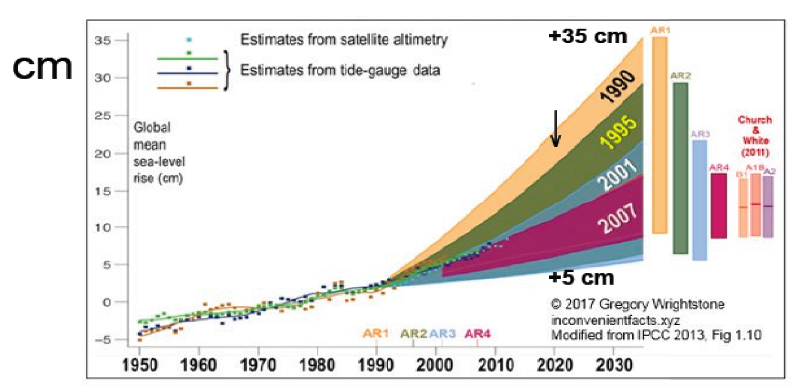
Figure 2.19 | Decadal global mean surface temperature (GMST) anomalies (white vertical lines in grey blocks) and their uncertainties (90% confidence intervals as grey

Uncertainty budget

An *uncertainty budget* is *not meant here* to be a tool allowing to (basically) reveal the uncertainty of the inference, e.g., through an inconsistent or imperfect "fitting" of the observed data behaviour, or through model uncertainty, as the only *two* components of uncertainty. Instead, it considers a fact that **no** *experimental datum* is *intrinsically* exempt from being affected by a larger number of components of uncertainty. The *above* components of uncertainty, instead, are *only the ones* that arise from the observed *dispersion* of data values—or the main causes why a *set* of data of different origin may have a dispersion in their observed trends. Any datum is invariably affected by a *larger* number of uncertainty components, which cannot be mitigated by any subsequent means.
In fact, each datum, when represented, should be decorated by its own uncertainty bar, as normally done in any metrological study.
The ensemble of the uncertainty components of a datum is called the "**uncertainty budget**" (UB), a *required feature in metrology*, but not so frequently used in other measurement fields.
One striking example has been found in the procedure used by IPCC and other Organisaitions for the purpose to treat the Surface-Ait Mean Temperature (SAMT) on the whole Earth.
Though the WMO has published a Guide about the evaluation of uncertainty (https://public.wmo.int/en/bulletin/communicating-forecast-uncertainty)—but not including an UB—then the communication to third parties of the collected SAMT values does *not* require the inclusion of uncertainty. These values usually include the first decimal digit, except in the METAR system where temperature is supplied as an integer number only.
However, a consistent scientific literature exists about the precision of the meteorological station, a largely standardised instrumental set, where the precision of the thermometer—which can be as good as ± 0.02 *C—is indicated to bring only a small contribution to the total uncertainty in the measurement of air, **± 0.5 *C** (e.g., P. Frank, Uncertainty in the global average surface temperature index: a representative lower limit, Energy & Environment, Vol. 21 No. 8, 2010)
In fact, the UBs consistently report the total value (rounded here) **± 0.5 *C** for each delivered measurement. That uncertainty is the "original" one of each SAMT, which cannot be mitigated by any elaboration procedure applied to the dataset, to obtain, according to the IPCC, what is intended to be the *worldwide* SAMT, whose published value and uncertainty of the last determination has been **(+1.12 ± 0.05) *C**, metrologically impossible. [2]
No uncertainty budget has been made available in that respect.
Similarly, most of the other climate parameters where found reported without their respective uncertainties. [3] For example, IPCC stated in the reported reference of Ch. 09 for the sea level rise: "The rate of melt water release from the Greenland and Antarctic ice sheets in response to climate change remains a major source of uncertainty in projections of sea level rise", without quantification.



SAMTestimate from NOAA



[2] In all instances, a fit of the annual change of the SAMT in the last 40 years brings to an **expanded uncertainty of ± 0.21 °C**. A fit of the SAMT database from HadCRUTS gave, with a 97,5% confidence, an uncertainty band of: 1850-1900 mean **0.5-0.8 °C**; 1980-2020 **0.2-0.3 °C**.That is not the total uncertainty but **only the component related to the data fit**.

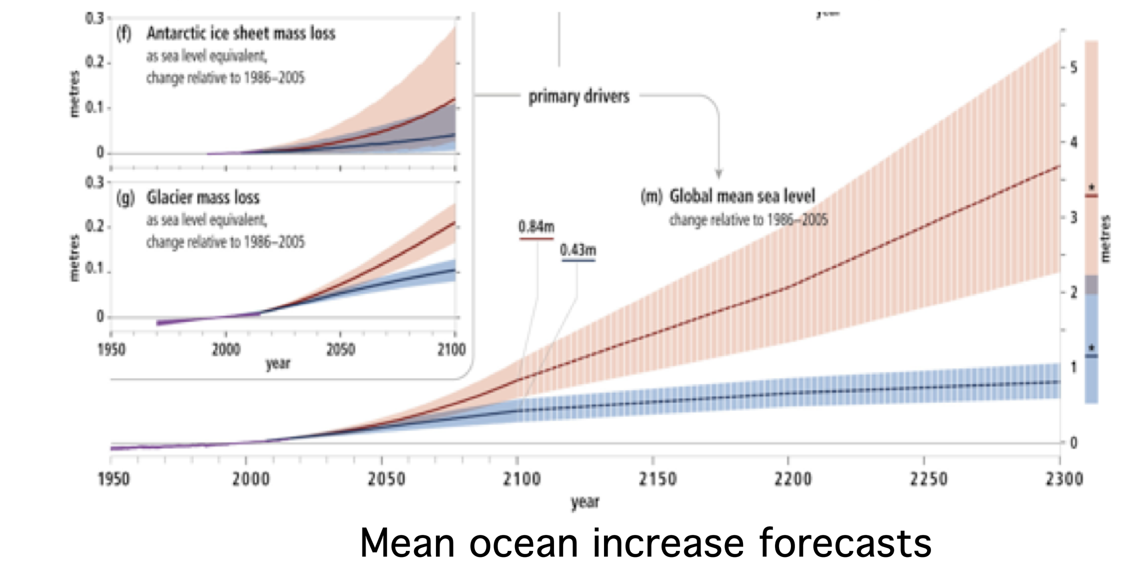
[3] The fact that they might be considered as **"consensus values"** does not justify the absence of the uncertainty indication. ***IPCC is not a normative Organisation***.

Effects of current knowledge uncertainty on the inference forecast for "*process-driven*" models

It is in the nature of any inference the fact that the wider/longer is the inference of the forecast the more critical is the quality/level of the current knowledge and of its uncertainty. Even the confidence on scientific method becomes more and more moderated: a good example is reported in the Figures. The forecast uncertainty is necessarily increasing, and the more rapidly the less reliable are the available data and for a shorter period.
In this respect, **the fact that, e.g., the true uncertainty of the SAMT is not (see above) ± 0.05 *C but, at best, ± 0.5 *C has certainly an adverse effect also on the reliability of the forecast of this parameter, also at short term.**
Similarly, for other important parameters like the rise of the sea level, the loss of ice in time, the level of CO_2 in the atmosphere ... … .

Effects of current knowledge uncertainty on the inference forecast for "*data-driven*" models

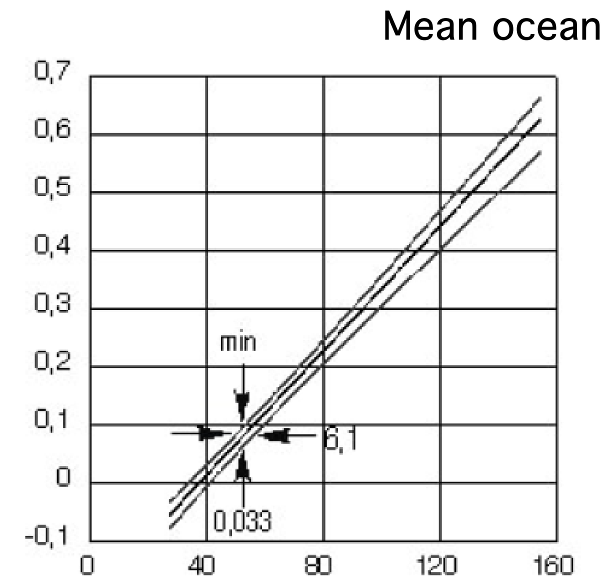
This type of model does not resort to theoretical bases, but is totally empirical in using only the available data and their distribution and trend, and the uncertainty of the latter, depending both on their quality, distribution and dispersion, as **treated via computer software and via computer-learning techniques**.
Non parametric well established methods exist to represent and qualify their characteristics, like the bootstrap or MonteCarlo ones. They allow to obtain an overall band on uncertainty of the dependent variable for a given confidence level over the full range of the independent variable where the current measurement exist.
Then, to explore inferences and their uncertainties, one resort to the characteristics of the specific "training"of the learning programme, instead of (only) on the simulation of future datasets.



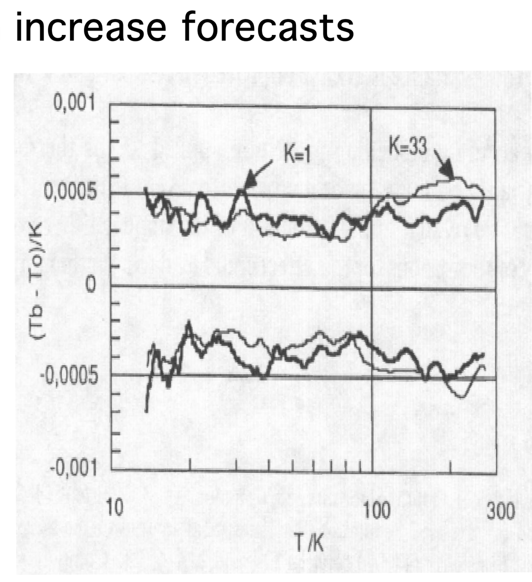
Mean ocean increase forecasts

Final remarks

A massive effort can be found in the scientific literature to deal with the *data uncertainty treatment* and the *inference uncertainty*.
However, in general, *data are treated as exact numbers*, and only the *dispersion* of their values is used for the above analyses.
In metrological terms, this means that only the *consistency* of the datasets is taken into account,
most often omitting *intrinsic* data uncertainty components, in some cases the most important components.
Thus, the conclusion of the analysis process is a **mere *deterministic function***, instead of a *band of variability* of the trend due to *data/model* uncertainty, and, when a band is shown, it only represents the effect of the *variety of used model functions*.
The consequence is that the inference uncertainty is generally affected by a (gross) under-evaluation,
especially when, as it happens in critical cases, the inferred period length exceed by far the one where (partial) knowledge is available.
The full *uncertainty budget* should be taken into account since the initial collation of the original data.
The "data-driven" techniques or hybrid methods seem to be more promising in providing more credible inferences.



Montecarlo



Bootstrap