Levels of reality in weather forecasting: the lesson by Richardson and von Neumann

Angelo VULPIANI

Dept. of Physics, University of Rome "Sapienza", Italy
Centro Interdisciplinare "B. Segre", Accademia dei Lincei

Why this talk?

Weather forecasting appears a very practical topic, nevertheless an analysis of its main aspects shows the presence of interesting conceptual topics:

* limits of extreme reductionism;
* limits of naive inductivism;
* relevance of old (apparently very far) classical issues;
* role of models at different scales;
* importance of the proper level of description.

Two opposite approaches to the prevision

EXTREME REDUCTIONISM

USE THE FIRST PRINCIPLES

Cold fronts are the way they are because of the properties of air and water vapour and so on which in turn are the way they are because of the principles of chemistry and physics.
We do not know the final laws of nature, but we know that they are not expressed in terms of cold fronts or thunderstorms.
(S. Weinberg, Nobel Prize in Physics)

NAIVE INDUCTIVISM

INFERE **ONLY** FROM THE DATA

Petabytes allow us to say: **“Correlation is enough”**. Therefore we can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.
The Data Deluge Makes the Scientific Method Obsolete
(C. Anderson, the prophet of the Big Data revolution)

We’ll see that both the above points of view do not work.

The basic idea of an inductive approach $\rightarrow$ BIG DATA

It seems natural to believe that

a *If a system behaves in a certain way, it will do so again*

b *From the same antecedents follow the same consequents*

Such claims are also supported by Biblical tradition:

What has been will be again, what has been done will be done again; there is **nothing new under the sun.**

(Qohelet's Book 1:9)

BIG DATA philosophy: forget the theory, now the data are enough.

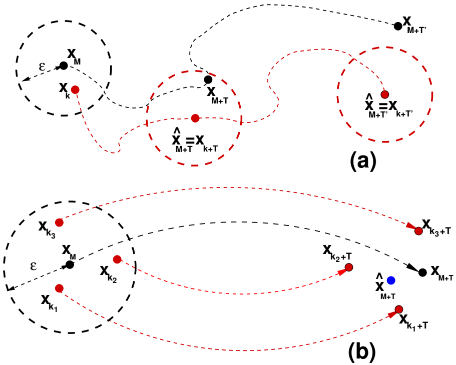
A formalisation of the idea "from the same antecedents follow the same consequents"

The method of the analogs

*- we know that the state of the system is given by a vector $\mathbf{x}$
*- we know the past of the system, i.e. a time series $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)$ where $\mathbf{x}_j = \mathbf{x}(j\Delta t)$
*- we want to predict the future, i.e. $\mathbf{x}_{M+t}$ for $t > 0$.

If the system is deterministic, in order to understand the future it is enough to look to the past for an "analog" i.e. a vector $\mathbf{x}_k$ with $k < M$ such that $|\mathbf{x}_k - \mathbf{x}_M| < \epsilon$, therefore, since "from the same antecedents follow the same consequents", we can "predict" the future at times $M + t > M$:

$$\mathbf{x}_{M+t} \simeq \mathbf{x}_{k+t}$$



A sketch of the method of the analogs

Conceptually everything sounds, however it is not so obvious at all that determinism holds, and it is easy to find and analog

It is a metaphysical doctrine that from the same antecedents follow the same consequents. ... But it is not of much use in a world like this, in which the same antecedents never again concur, and nothing ever happens twice. ... The physical axiom which has a somewhat similar aspect is "That from like antecedents follow like consequents."

(**James Clerk Maxwell**)

The forecast is based on the supposition that what the atmosphere did then, it will do again now.....
The " Nautical Almanac", that marvel of accurate forecasting, is not based on the principle that astronomical history repeats itself in the aggregate. It would be safe to say that a particular disposition of stars, planets and satellites never occurs twice. Why then should we expect a present weather map to be exactly represented in a catalogue of past weather?

(**Lewis Fry Richardson**)

Attempts to the forecasting using the analogs

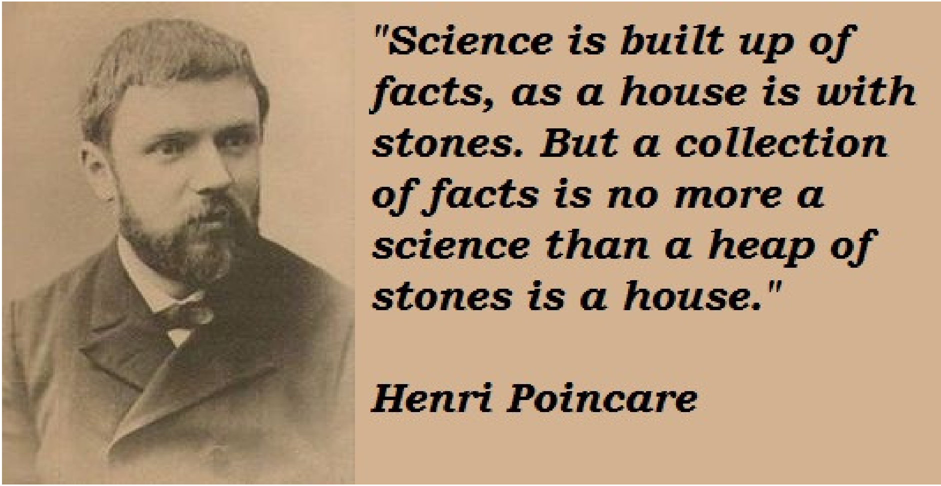
Lorenz tried to use the meteorological charts of the past to perform a weather forecasting. Applying the method of the analogs he realised that the intuition of Richardson is correct.

In practice, this procedure may be expected to fail, because of the high probability that no truly good analogues will be found within the recorded history of the atmosphere.

Also the methods used in finance (where the state $\mathbf{x}$ is not known) for the prediction basically, are (stochastic) versions of the analogs.

Now we are in the DATA DELUGE age.
Can we hope to success using just data?

Big Data: cornucopia or Pandora's box?



Looking back to an apparently very far topic

The Poincaré recurrence theorem
In a deterministic system with a bounded phase space, after a certain time, the system must be close to its initial state

Such a theorem had a great historical relevance in the strong debate, at the end of the 19-th century, between Boltzmann and Zermelo on the irreversibility.
Boltzmann had been able to show, with probabilistic arguments, that in a system with $N \gg 1$ particles the recurrence is not a real problem: the return time is very large

$$T_R \sim \tau_0 C^N$$

where τ_0 is a characteristic time and $C > 1$, in a macroscopic system ($N \sim 10^{20} - 10^{25}$), T_R is gigantic, much larger that the age of the universe.

A simple, but important, result from the ergodic theory

The intuition of Boltzmann had been formalised by the **Kac Lemma**
In an ergodic ergodic system the average return time $\langle \tau(A) \rangle$ *in a set* A *is*

$$\langle \tau(A) \rangle = \frac{\tau_0}{P(A)}$$

where $P(A)$ *is the probability to be in* A*.*
Consider a system of linear sizes $O(\epsilon)$, therefore $P(A) \sim (\frac{\epsilon}{L})^D$ so

$$\langle \tau(A) \rangle \sim \tau_0 \left(\frac{L}{\epsilon} \right)^D$$

where L is the excursion of each component of the vector describing the state and D the attractor's **dimension**.

Consequences of the Kac Lemma

Irreversibility The Boltzmann's intuition was correct.
Since $D \sim N \gg 1$, macroscopic irreversibility is not in disagreement with the Poincaré recurrence theorem, the return time is too large:

$$\tau_0 \left(\frac{L}{\epsilon} \right)^D$$

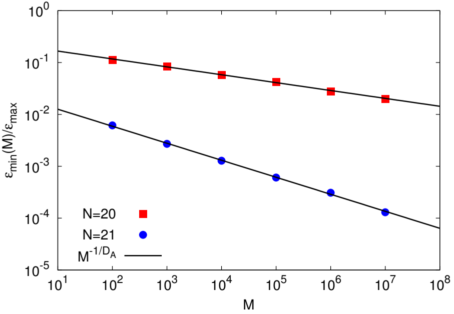
Forecasting In order to find an analog, the size M of the time series must be, at least, of the same order of the recurrence time:

$$M_{min} \sim \frac{\tau_0}{\Delta t} \left(\frac{L}{\epsilon} \right)^D$$

Since in the atmosphere D is not small, Lorenz had no chance to find an analog. Even with a limited precision, say 5%, i.e. $L/\epsilon = 20$, one has that, if D is large, say 6 or 7 it is pretty impossible to find an analog.

A toy model (proposed by Lorenz) for the weather helps to understand the difficulty

$$\frac{dx_n}{dt} = x_{n-1}(x_{n+1} - x_{n-2}) - x_n + F \ \ , \ \ n = 1, 2, ..., N$$



The relative precision of the best analog as function of the size of the time series. Two systems with $F = 5$, for $N = 21$ one has $D \simeq 3.1$ (circles), for $N = 20$, $D \simeq 6.6$ (squares).

Lewis F. Richardson (1881- 1953), the great visionary

Weather Prediction by Numerical Process

In his seminal book Richardson proposed to use the equations regulating the evolution of the atmosphere.

The atmosphere evolves according to the equations of hydrodynamics (for the fields describing velocity $\mathbf{u}$, density ρ, pressure p, water percentage s, and temperature T) and the thermodynamics giving the relation (equation of state) among ρ, T, s and p.
So, by knowing the present state of the atmosphere, we can solve seven partial differential equations to obtain– at least in principle– a weather forecast. Of course, these equations cannot by solved by pen and paper, so a numerical solution is the only option.

The first heroic numerical attempt

The initial conditions used by Richardson consisted of a record of the weather charts observed in Northern Europe at 4 A.M. on 20 May 1910 during an international balloon day.
The numerical work by Richardson was long, taxing and wearisome: it has been estimated that, in the course of two years **he worked for at least one thousand hours**, computing by hand and with some rudimentary computing machine. The result, giving **a six-hour forecast**, was quite disappointing.

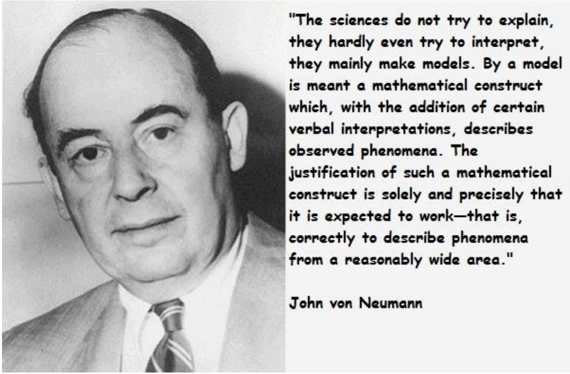
Richardson correctly understood that *the scheme is complicated because the atmosphere is complicated.*
Nevertheless, he was moderately optimistic in his conclusive remarks: *perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances. ... But that is a dream.*

The failure is because the equations proposed by Richardson are too accurate!

The original Richardson's attempt, based on the first principle, is, somehow, a form of reductionism.

The realisation of Richardson's dream had to wait until the 1950s.
Instead of the "obvious" use of the first principles, it has been necessary to adopt another approach which include the development of three "ingredients", all far from trivial
a) **effective equations**;
b) fast numerical algorithms;
c) computers suitable for numerical calculations.

John Von Neumann (1903- 1957), a pragmatic scientist



The Meteorological Project

For the weather forecasting, and more general, in any "complex" problem, it is necessary to understand which aspects have to be taken into account and which ones can be ignored.

To develop the skill of correct thinking is in the first place to learn what you have to disregard. In order to go on, you have to know what to leave out: this is the essence of effective thinking.
(**Kurt Gödel**)

Fast phenomena, e.g. waves, are not especially interesting for weather forecasting, but they influence the slow variables, so they have to be somehow accounted for. The way to solve the problem was found by Charney, von Neumann and colleagues in the 1940s- 1950s, within the Meteorological Project at the Institute for Advanced Study, in Princeton. The project involved scientists from different fields: mathematicians, experts in meteorology, engineering, and computer science.

The effective equations

Almost all the interesting dynamic problems in science and engineering are characterised by the presence of more than one significant scale, i.e. there is a **variety of degrees of freedom with very different time scale**, e.g.
***- protein folding**: the time scale of vibration of covalent bonds is $O(10^{-5})s$, the folding time for proteins may be of the order of seconds.
***- climate**: the characteristic times of the involved processes vary from days (for the atmosphere) to $O(10^3)yr$ (for the deep ocean and ice shields).

The necessity of treating the "slow dynamics" in terms of effective equations is both practical (even modern supercomputers are not able to simulate all the relevant scales involved in certain difficult problems) and conceptual: effective equations are able to catch some general features and to reveal dominant ingredients which can remain hidden in the detailed description.

The simplest case: only two characteristic times

Consider a system whose state is given by $\mathbf{X} = (\mathbf{X}_f, \mathbf{X}_s)$ where $\mathbf{X}_f$ and $\mathbf{X}_s$ are the fast and slow components.

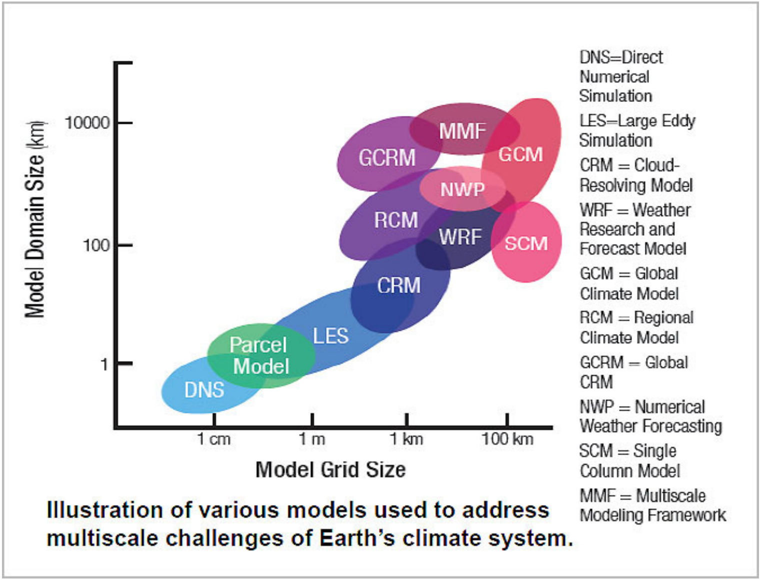
$$\frac{d\mathbf{X}_s}{dt} = \frac{1}{\tau_s} F_s(\mathbf{X}_f, \mathbf{X}_s)$$

$$\frac{d\mathbf{X}_f}{dt} = \frac{1}{\tau_f} F_f(\mathbf{X}_f, \mathbf{X}_s)$$

with $\tau_f \ll \tau_s$.
The aim is to derive an "effective" equation only for $\mathbf{X}_s$:

$$\frac{d\mathbf{X}_s}{dt} = \frac{1}{\tau_s} F_{eff}(\mathbf{X}_s) \ .$$

This idea has been successfully used by Langevin for the diffusion of colloidal particles in a fluid (Brownian motion).



For practical purposes the equations used by Richardson are appropriate just for spatial scales smaller than $O(10)$ km.

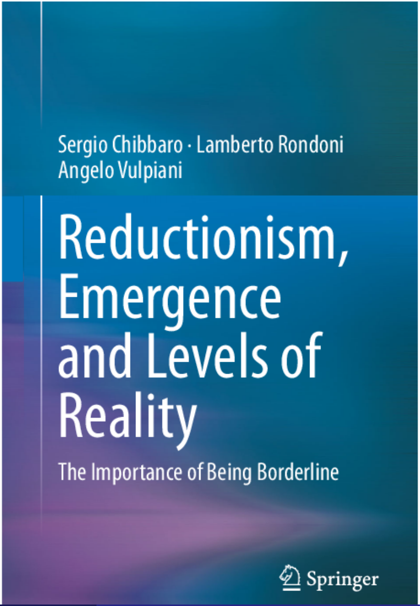
About the effective equations

a From a computational point of view: it is possible to use larger Δt and Δx in the numerical integration;

b Their description of the slow dynamics make it possible to detect the most important factors, which on the contrary remain hidden in the detailed description given by the original equations.

c They are not mere approximations of the original equations, typically emergent features appear.

Levels of reality: an advertisement



Examples of levels of reality

Statistical Mechanics
I- microscopic level, Γ- space description (Liouville equation);
II- microscopic level, μ- space description (Boltzmann equation);
III- mesoscopic level, μ- space description but at "large scale" (Fokker–Planck equation);
IV- macroscopic level, fluidodynamics description (Navier– Stokes equation, Fourier law, . . .).
Climate
I- molecular level
II- fluid dynamics
III- quasi-geostrofic equations
IV- effective equations

The crossing from one level of description to another is rather delicate, it is determined by a coarse- graining and/or a projection procedure with a "loss of information".

Model from data?

If it is not possible to use models "derived" from some well based theory (e.g. classical or quantum mechanics) it seems natural to use an inductive approach.

The building of model from data

In the case (very rare) we know the vector $\mathbf{x}_t$ describing the state of the system, at least in principle one can adopt the method of the analogs looking back in the past and then build a map

$$\mathbf{x}_{t+1} = \mathbf{G}(\mathbf{x}_t)$$

where the shape of $\mathbf{G}$ can be obtained with some fitting/optimization procedure.

The troubles

Trouble 1 Even in the (lucky) case we know the proper variables $\mathbf{x}_t$ if the dimension is larger that 5 or 6 it is pretty impossible to find analogs, there the protocol collapses

Trouble 2 *Typically we do not know the proper variables*
Such rather serious difficulty is well known, for instance in statistical physics:

How do you know you have taken enough variables, for it to be Markovian? [Onsager and Machlup]

The hidden worry of thermodynamics is: we do not know how many coordinates or forces are necessary to completely specify an equilibrium state. [Ma]

We have a nice mathematical result,
BUT IT CANNOT SOLVE ALL THE PROBLEMS...

Takens gave an important contribution to the understanding the problem in the case we do not know the proper variables:
From the study of a time series $\{u_1, .., u_M\}$, *where* u_j *is an observable sampled at the discrete times* $j\Delta t$, *it is possible (if we know that the system is deterministic and is described by a finite dimensional vector, and* M *is large enough) to determine the proper variable* $\mathbf{x}$.

Unfortunately, at practical level, the method has rather severe limitations:
A) It works only if we know *a priori* that the system is deterministic;
B) The protocol fails if the dimension of the attractor is large enough (say more than 5 or 6).

No evidence at all of a blind inductive approach to build a model

Conclusions and Remarks

The idea (dream) to avoid the theory and use only data, is too naive. Because of the Kac's lemma, the BIG DATA approach can work only for very low dimensional systems.

Old topics can be relevant even in modern practical issues: e.g. the Poincaré recurrence theorem (and Kac's lemma) for the analogs.

It is true that *the final laws of nature are not expressed in terms of cold fronts or thunderstorms*, however the unique way to understand the atmosphere is to write down **effective equations** for the cold fronts.

The dream to build models just from data cannot work if the dimensionality of the problem if large enough ($D > 5$ or 6).

Some References

* L.F. Richardson *Weather Prediction by Numerical Process* (Cambridge University Press, 1922)
* J.G. Charney *On a physical basis for numerical prediction of large-scale motions in the atmosphere* J. Meteor. **6**, 371 (1949)
* J.G. Charney, R. Fjortoft and J. von Neumann, *Numerical integration of the barotropic vorticity equation* Tellus **2**, 237 (1950)
* M. Kac *Probability and Related Topics in Physical Sciences* (Am. Math. Soc. 1957)
* E.N. Lorenz *Three approaches to atmospheric predictability* Bull. Am. Met. Soc. **50**, 345 (1969)
* E. N. Lorenz *Atmospheric predictability as revealed by naturally occurring analogues* J. Atmos. Sci. **26**, 636 (1969).
* E.N. Lorenz *The essence of chaos* (UCL Press, 1995)
* A.S. Weigend and N.A. Gershenfeld (Editors) *Time series prediction* (Addison-Wesley Pub., 1994)

* H. Kantz and T. Schreiber *Nonlinear time series analysis* (Cambridge University Press, 1997)
* A.D. Dalmedico *History and epistemology of models: Meteorology as a case study* Archive for history of exact sciences **55**, 395 (2001)
* O. Darrigol *Worlds of Flow: A history of hydrodynamics from the Bernoullis to Prandtl* (OUP Oxford, 2005)
* P. Lynch *The Emergence of Numerical Weather Prediction: Richardson's Dream* (Cambridge University Press, 2006)
* C. Anderson *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* http://www.wired.com/2008/06/pb-theory/
* P. Castiglione, M. Falcioni, A. Lesne and A. Vulpiani *Chaos and Coarse Graining in Statistical Mechanics* (Cambridge University Press, 2008)
* F.Cecconi, M. Cencini, M. Falcioni and A.Vulpiani *The prediction of future from the past: an old problem from a modern perspective* Am. J. Phys. **80**, 1001 (2012)
* S. Chibbaro, L. Rondoni and A.Vulpiani *Reductionism, Emergence and Levels of Reality* (Springer-Verlag, 2014)
* C.S. Calude and G.Longo *The Deluge of Spurious Correlations in Big Data* Found. Sci., March 2016