

# Big Data, Weathermen and Economists: Never fall in love with a model

Miia Parnaudeau<sup>1</sup>

<sup>1</sup>Affiliation not available

April 17, 2023



# Big Data, Weathermen and Economists: Never fall in love with a model

MIIA PARNAUDEAU

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:  
[mia.parnaudeau@essca.fr](mailto:mia.parnaudeau@essca.fr)

DATE RECEIVED:  
June 10, 2015

DOI:  
10.15200/winn.143073.30255

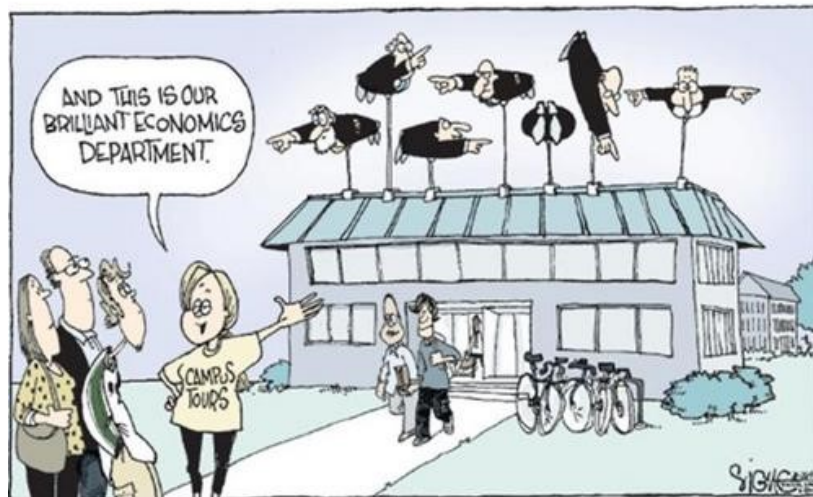
ARCHIVED:  
May 04, 2015

CITATION:  
Mia Parnaudeau, Big Data, Weathermen and Economists: Never fall in love with a model, *The Winnower* 2:e143073.30255, 2015, DOI: 10.15200/winn.143073.30255

© Parnaudeau This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



The overabundance of data which now characterizes the work environment of meteorologists, data scientists and risk managers imposes a methods renewal<sup>[1]</sup>[meteosensibilite.com/gestion-du-risque-meteo/big-data-correlation-causalite-3098#more](http://meteosensibilite.com/gestion-du-risque-meteo/big-data-correlation-causalite-3098#more)\n3098. And so Machine Learning entered into weather modelisation. Because they have chosen to work with non-linear models rather than linear ones, the weather specialists are presented as quick to adopt new techniques, more suited to the realities they are trying to capture. This made even some say that weathermen had much to teach to economists<sup>[2]</sup>[www.forbes.com/sites/stevekeen/2015/04/16/you-do-need-a-weatherman/](http://www.forbes.com/sites/stevekeen/2015/04/16/you-do-need-a-weatherman/), because they fail to abandon their old econometric 'recipes'.



Does Machine learning

really solve all the problems at once? Are Linear Models really outdated?

HELLO WORLD!<sup>[3]</sup>[HELLOWORLDCOLLECTION.DE/](http://HELLOWORLDCOLLECTION.DE/)

The first model of a learning machine was developed by Frank Rosenblatt in 1957 and was called the 'perceptron'. Rosenblatt described the model as a program for computers and demonstrated with simple experiments that this model can be generalized<sup>[4]</sup>[www.springer.com/us/book/9780387987804](http://www.springer.com/us/book/9780387987804). A Learning Machine allows computers to learn from the data. This theory has then flourished with the work of the mathematicians Vladimir Vapnik and Alexey Chervonenkis. Machine learning is a branch of artificial intelligence. The intelligence is built by referring to examples.

Something important is changing in how we as a society use computers to mine data<sup>[5]</sup>[www.cs.cmu.edu/~tom/](http://www.cs.cmu.edu/~tom/). Machine learning algorithms have helped to reveal trends and patterns that are too subtle for humans to detect. Machine learning is already an incredibly powerful tool that can help to solve really hard classification problems<sup>[6]</sup>[www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/](http://www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/). However, if its predictive power is real, it is not possible to interpret the reasons supporting the results obtained. As David Karger, professor of computer science at the MIT recently noticed, the algorithm 'just works like magic', and 'you don't really know \*why\* that's the answer'.

### **WEATHER ANOMALIES AND MACHINE LEARNING**

For Hamada Badr, Benjamin Zaitchik and Seth Guikema<sup>[7]</sup>Badr H.S., Zaitchik B.F., and Guikema S.D., (2014), "Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel", *Journal of Applied meteorology and climatology*, 53, 614- 636, Machine Learning clearly matters in the study of weather anomalies. While arguing that conventional statistical modeling techniques as generalized linear models (GLMs) have demonstrated some skill, they rather propose the use of an artificial neural network (ANN) machine-learning algorithm.

With this algorithm, they predict summertime rainfall anomalies in the Sahel region of Africa. The ANN has there a better accuracy and fit. Its main skill, compared to other kinds of models, is its ability to capture nonlinear influences. A study on Sahel rainfalls in summer reminds me necessarily the words of George Carlin – e.g. weather forecast for tonight: dark-. But more seriously, do these approaches really sign the end of old econometric 'recipes'?

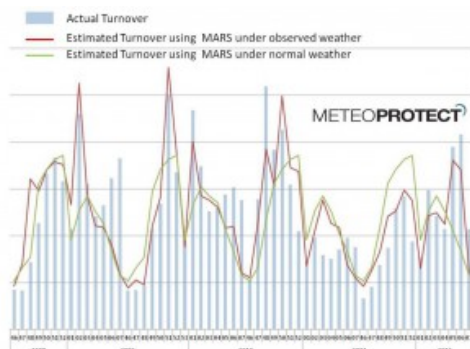
### **A STORM IN A TEA-CUP**

I rather feel like this is a storm in a tea-cup. The recent blog posts of Arthur Charpentier<sup>[8]</sup>[freakonometrics.hypotheses.org/19424](http://freakonometrics.hypotheses.org/19424) are worth a read to show why. In response to the common criticisms addressed by Machine learning specialists, arguing that it's easy to find a machine learning 'model' that will always beat any kind of Generalized Linear Model, he demonstrated quite clearly the contrary. It is still possible to rely on standard econometric models.

In the case of the weather, a good example comes through the work of Max Little, Patrick McSharry and James Taylor<sup>[9]</sup>Little, M.A., McSharry P.E., Taylor J.W., (2009), "Generalized Linear Models for Site Specific Density Forecasting of U.K. Daily Rainfall", *Monthly Weather Review*, 137, 1029-45. In the U.K., site-specific probabilities of rainfall are needed to price insurance contracts. GLMs perform best in this specific case and are consequently highly recommended for forecasting purposes. Everything comes through a detailed study of the interaction between the autocorrelation properties of the data and the structure of the models tested.

So, there are good old econometric 'recipes' still used. And they still have much more to tell us as we get further into the foundations of what drives the links within the data. Linear models still have some words to say!

Old recipes work even when taking into account non-linearities. Another good example of this comes from a study provided by MeteoProtect to a clothing retailer in 2012. This study used notably a multivariate adaptive regression splines model (MARS). These models are linear, but they automatically model non-linearities and interactions between variables. Here again, we are able to get into the interpretation of what characterizes the Data Generating Process.



The quality of the adjustment using a MARS model was the best, and so it was retained. But what these models tell us here is much more than only predictive. If we compare the estimated turnover using MARS and observed weather (red curve), the estimated turnover using MARS and normal weather (green curve) to the turnover (blue), one can firstly see how much the actual turnover is weather-sensitive.

Weather risk management earns a lot from this kind of solution. And, beyond the hedging solutions that can be provided on this basis, the CEO gets informed on many more elements. Once the periods of high sensitivity to weather are identified, situations where the turnover has not been affected by the weather (being normal or not) can be addressed. An example of this can be seen in the beginning of the year 2011. The disconnection of the actual turnover (blue) from the estimated ones (green and red) using weather variables informs us on the fact that on this specific period, other variables have had an impact on the activity. In understanding what really drives these movements, a CEO will be better able to pilot his performance over the time. This information is hidden within the Data Generating Process.

Old recipes have still their utility. It's not about opposing linear to non-linear approaches, machine learning to econometrics, meteorologists to economists and so on... MARS models have proven to be unsatisfactory in other situations, and alternative approaches have been used, once experimented, by **Meteo Protect**. This is another story.

The real story lies behind the data. This has been understood a long time ago both by meteorologists and economists. It is a question of methodology, as it has been explained in 1987 by Adrian Pagan<sup>[10]</sup> Pagan, A.R., (1987), "Three Econometric Methodologies: A Critical Appraisal", *Journal of Economic Surveys*, 1(1), January, 3-24: first, the methodology should provide a set of principles to guide work in all its facets; second, by codifying this body of knowledge it should greatly facilitate the transmission of such knowledge; finally, a style of reporting should naturally arise from the methodology that is informative, succinct, and readily understood.

The main lesson that stems from this, is that you should never – ever – fall in love with a model.

#### References

1. ↑ [meteosensibilite.com/gestion-du-risque-meteo/big-data-correlation-causalite-3098#more-\n3098](http://meteosensibilite.com/gestion-du-risque-meteo/big-data-correlation-causalite-3098#more-\n3098)
2. ↑ [www.forbes.com/sites/stevekeen/2015/04/16/you-do-need-a-weatherman/](http://www.forbes.com/sites/stevekeen/2015/04/16/you-do-need-a-weatherman/)
3. ↑ [helloworldcollection.de/](http://helloworldcollection.de/)
4. ↑ [www.springer.com/us/book/9780387987804](http://www.springer.com/us/book/9780387987804)
5. ↑ [www.cs.cmu.edu/~tom/](http://www.cs.cmu.edu/~tom/)
6. ↑ [www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/](http://www.forbes.com/sites/quora/2015/02/12/what-is-the-future-of-machine-learning/)
7. ↑ Badr H.S., Zaitchik B.F., and Guikema S.D., (2014), "Application of statistical models to the prediction of seasonal rainfall anomalies over the Sahel", *Journal of Applied meteorology and climatology*, 53, 614- 636
8. ↑ [freakonometrics.hypotheses.org/19424](http://freakonometrics.hypotheses.org/19424)
9. ↑ Little, M.A., McSharry P.E., Taylor J.W., (2009), "Generalized Linear Models for Site Specific Density Forecasting of U.K. Daily Rainfall", *Monthly Weather Review*, 137, 1029-45
10. ↑ Pagan, A.R., (1987), "Three Econometric Methodologies: A Critical Appraisal", *Journal of*

Economic Surveys, 1(1), January, 3-24