



# The importance of agent-based modelling for official statistics

Frank P. Pijpers

November 2021

# CBS = Statistics Netherlands

The national statistical institute for the Netherlands  
(equivalent to ONS in the UK or DeStatis in Germany e.g.)

The screenshot shows the StatLine website interface. At the top, there is a search bar and navigation links for Themes, Recent, and Help. The main title is "Renewable electricity; production and capacity", with a note that it was last changed on 12 July 2021. A helpful message states: "Variables can be dragged to the header, rows or columns of the table. In the header only one item of a variable can be selected." Below this, a table is displayed with a dropdown menu for "Topic" set to "Production of electricity". The table has several columns: "Energy sources / techniques", "Periods", "Production of electricity" (with sub-columns for "Gross production with normalisation", "Production without normalisation", and "Net production of electricity"), "Production of electricity relative" (with sub-columns for "Gross production with normalisation" and "Production without normalisation"), "Installed installations" (with sub-columns for "number" and "megawatt"), and "Elect capacity er". The data rows show values for "Total renewable energy sources", "Hydro power", "Wind energy, total", "Wind energy on shore", and "Wind energy of shore" for the years 2019 and 2020.

Energy sources / techniques	Periods	Production of electricity			Production of electricity relative		Installed installations		Elect capacity er
		Gross production with normalisation	Production without normalisation	Net production of electricity	Gross production with normalisation	Production without normalisation	number	megawatt	
Total renewable energy sources	2019 <sup>est</sup>	22,213	22,732	21,978	18.22	18.65	18.56	.	
	2020 <sup>est</sup>	31,235	32,315	31,364	26.03	26.93	26.90	.	
Hydro power	2019 <sup>est</sup>	93	74	74	0.08	0.06	0.06	7	
	2020 <sup>est</sup>	90	46	46	0.07	0.04	0.04	.	
Wind energy, total	2019 <sup>est</sup>	10,774	11,508	11,508	8.84	9.44	9.72	2,321	
	2020 <sup>est</sup>	15,945	15,339	15,339	11.62	12.78	13.16	2,606	
Wind energy on shore	2019 <sup>est</sup>	7,429	7,935	7,935	6.09	6.51	6.70	2,032	
	2020 <sup>est</sup>	8,960	9,856	9,856	7.47	8.21	8.45	2,144	
Wind energy of shore	2019 <sup>est</sup>	3,345	3,573	3,573	2.74	2.93	3.02	289	
	2020 <sup>est</sup>	4,985	5,484	5,484	4.15	4.67	4.70	443	



# Why should Statistics Netherlands (CBS) do modelling?

‘Classical’ argument: CBS / official statistics does not use models  
This is nonsense: every estimator implicitly presupposes a model

2<sup>nd</sup> inappropriate argument: CBS does not do predictions which is what models are for: ‘territorial claims’ from the time of Tinbergen

Before a model is ‘quality certified’ for the use of making predictions it has to be calibrated / validated, e.g. using the (recent) past.

For the social and economic domains no one else has the data, so we do have (the responsibility to) build up the necessary expertise so that we can do this calibration work.

# Why agent based models (ABMs)

Characteristic for economy and society:

- population
- interactions

There are 17+M inhabitants in NL. The collective effect of individual interactions is emergent behaviour of the system: trends, social structures, segregation & integration.

Interaction at group level provides a biased view, and ignores the skewed distributions of a variety of properties of individuals within a group.

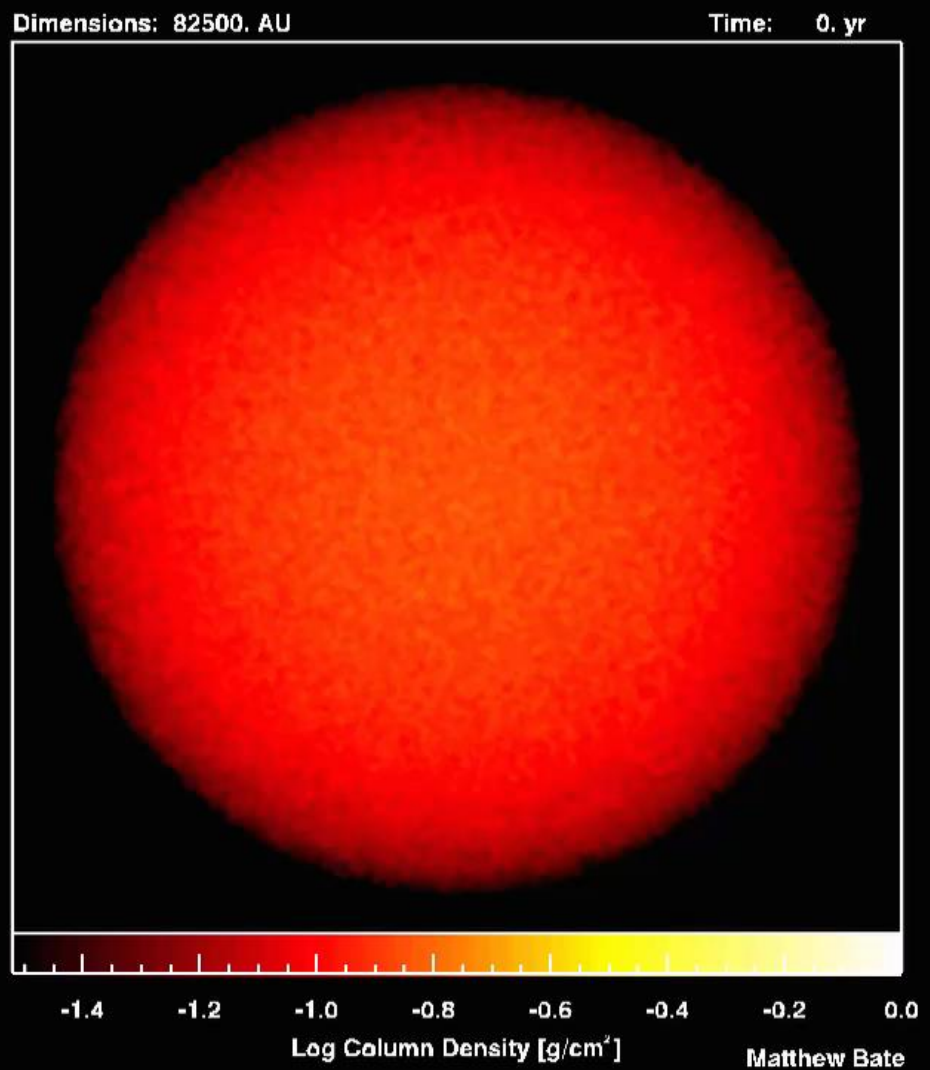


A demonstration of a ‘many body’ system with a simple interaction rule (gravity).

Even when starting from an almost homogeneous situation an enormous palette of structures is produced: emergent behaviour.

In our application 3D coordinates are replaced by distances along social axes: family/ neighbours/ colleagues.

Asymmetric interactions will very quickly lead to much ‘richer’ emergent behaviour.



# Where do we want to head?

## Shorter term:

- Working proof of concept:
  - using CBS data; provide all agents with characteristic properties and set up appropriate interaction rules to run simulations
  - develop analysis tools and visualisations for the results
- Research mechanisms that feed inequality resp. prevent it : education, labour market, legal frameworks, crises & other (external) factors etc.
  - starting from the '50s or '90s, simulate forward in time (data completeness)

# Where do we want to head?

## Longer term

- Fast & general purpose socio-economic simulation tool that can be offered as a service to planning offices en universities
  - validation and calibration of model parameters: on-the-fly or as pre-simulation research
- Broadening of application to other countries (e.g. via Eurostat) and other subject areas (climate adaptation, social polarisation, ...)

