
BIOFUELS' UNBALANCED EQUATIONS: MISLEADING STATISTICS, NETWORKED KNOWLEDGE AND MEASURED PARAMETERS

Part I: Evolution of globalised soil, land and terrain databases¹

KATE B. SHOWERS

Centre for World Environmental History
University of Sussex

Abstract

Biofuels were central to the European Union's early twenty-first-century responses to Kyoto-mandated carbon emissions reduction. This paper, the first part of three, traces the evolution of eight globalised soil, land and terrain databases and models underpinning global biofuels research. Central to the discussion are examination of the meaning and power of numbers; the significance of homogenising and harmonising data; and the decline of data quality inherent in derivation. The relevance of globalised model-based analysis for practical policy formulation and local implementation are discussed.

Keywords: European Union energy policy, data quality, GLASOD, GAEZ, World Soil Map, potential arable land

Introduction

Major strategies to mitigate climate change assume widespread and increased use of electricity as a substitute for fossil fuels. Ignored is the fact that electricity is a *form*, not a *source*, of energy. When subjected to analysis from environmental history perspectives, the extent to which proposals for electricity generation can be

¹ Ideas first presented in an invited paper at the Gordon Research Conference on Industrial Ecology, Les Diablerets, Switzerland, 17–22 June 2012, and further developed for an invited paper CAS Herbsttagung 2015, Universität für Bodenkultur, BOKU, Vienna, Austria, 20 November 2015, available on podcast at www.youtube.com/watch?v=tde9mj9g-VI. The author would like to thank the anonymous reviewers for constructive comments.

considered to be 'green', renewable or sustainable is brought into question. This is clearly demonstrated when the design and implementation of the European Union's (EU) biofuels policy as a response to its Kyoto-mandated responsibilities is analysed over several decades.

The cultivation of agricultural crops as biofuels feedstocks (agrofuels) for use in transportation and electricity generation seemed an ideal European response to the threat of increasing carbon emissions linked to global climate change. Policy enthusiasm collided with agronomic possibility when landscapes proved to have yield limitations; extensive monocrops were shown to be unsustainable; and biofuels feedstock cultivation increased competition for land among food, feed/forage crops, forestry and the environment. Expansion of the proposed production area to the entire world did not resolve these conflicts.² However, renewed optimism came with a change of scale: globalised databases and models using them assured the existence of ample land. Third-generation biofuels (non-agricultural crops grown on hypothesised non-agricultural land) would circumvent the food versus fuel competition presented by agrofuels and second-generation (non-food crop) biofuels. The search for cultivable land was guided by models incorporating a few plant growth parameters and operated with globalised soil, land or terrain databases. The results are expressed in units of grid cell rather than geographical area.

In these models, soils are recognised only as mediums for plant growth. They are defined by a limited number of physical and chemical aspects identified as benefiting or hindering specific crops. This represents not only a gross simplification of world agricultural and livestock production, but, more fundamentally, an under estimation of the significance of soils. Whether from the scale of soil particle composition or that of regionally expressed properties, soils are complicated systems essential to the structure, function and reproduction of flora and fauna, biome processes and earth-atmosphere dynamics. Identifying soil for agricultural use is more complex, and disrupting uncultivated soils is more detrimental than modellers and policymakers imagine.

Given the change of scale and increase in mathematical manipulations in database construction, researchers should practice 'data literacy'. The idea of data literacy follows from the concept of media literacy, which assists consumers to deconstruct advertising and media in general. Data literacy could serve as a tool for assessing the quality of globalised statistics and analyses generated by models designed to operate with them. Before any modelling activity, fundamental questions should be asked about data collection, data set and database creation, and model construction in order to have a clear understanding of the validity of the numbers generated.

2 Kate B. Showers, 'Land Use from Below: Biofuels, Urbanization and Sustainable Soil Management in Europe and Africa', in *The Challenge of Sustaining Soils: Natural and Social Ramifications of Biomass Production in a Changing World*, ed. Verena Winiwarter and Martin Gerzabek (Vienna: Austrian Academy of Sciences, 2012).

As a contribution to debates about the practicality of biofuels cultivation, this paper presents an exercise in data literacy in the context of an environmental history of globalised soil, land and terrain databases and models. The paper will begin with a reflection on the meaning and power of numbers as well as on data quality. Discussion of the evolution of the databases and models will serve as a base for subsequent examination of the logic of data collection and manipulation, database development and the origins of models used in global crop production analysis.

Data

Data begin as numbers generated by counting or measurement to provide a quantitative description. Numbers become data when aggregated, and acquire meaning through manipulation. In grammatical analysis, numbers are adjectives modifying nouns. On their own, they are meaningless, but they have power because, unlike non-quantitative adjectives such as 'little', 'fast' or 'heavy', numbers can be used in calculations, enabling generalisations from individual observations.

Mathematicians developed statistical analysis to ensure accurate quantitative descriptions of groups and to provide confidence in projections. Statistics' mathematical operations follow clearly specified rules to preserve the integrity of essential assumptions and the validity of conclusions drawn. A central concern is reducing variability to enable extrapolation of general statements to clearly specified larger populations. Calculating averages, depicted in Figure 1, is a fundamental tool for minimising difference or homogenising data.

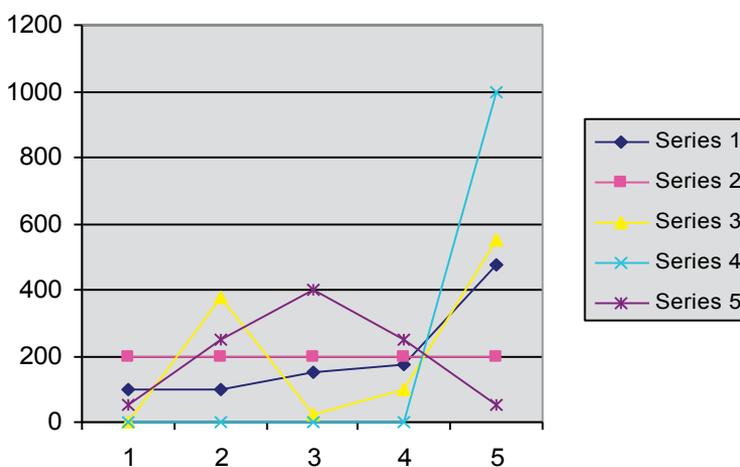


Figure 1: Five series of observations with an average of 200.

Each of the five sets of observations depicted has an average value of 200, but only one is described precisely by 200—that in which all data points equal 200. However, the other data sets representing different amounts and kinds of variability are accurately described as being, on average, 200. If each of the points on the graph themselves represent averaged data collections, the graphed value would represent a further masking of variability, and a greater distance from the diversity represented in the original observations. Other accepted methods for reducing variability include removing outlying data points from data sets for subject matter reasons, and mathematical techniques for data smoothing that fit numbers to expected (theorised) curves.

As discussed by Erik Reinert,³ the transformation of observations into quantitative description at ever-higher levels of generalisation can be understood to move along a continuum from abstraction to simplification.⁴ In abstraction, taxonomies of detailed information support construction of a representative group. This traceable system of relationships provides an underlying and stable base for theoretical development. In contrast, simplification involves the removal of context and the denial of the significance of difference. Simplification eliminates elements of quality and interdependence, focusing, instead, upon quantity. In this way, numbers themselves become more important than what they once represented. The highest levels of simplification achieve what Reinert calls ‘practical irrelevance’.⁵ It could be argued that, stripped of contextual meaning, what remains is the embedded power of the world view that informed each decision to exclude, or, more simply, deny diversity while asserting homogeneity. This continuum should be kept in mind when creating or using databases and statistics at global scales.

Power and data quality

From data collection to database construction, researchers are confronted by a series of decisions about what to include and what to exclude. These choices will be informed by the researchers’ personal and professional value systems. While averaging makes exclusion a technical exercise, engaging in data smoothing—and certainly eliminating outlier data points—is more subjective. Social scientists have identified the ability to exclude as a form of power.⁶ Data set and database constructors thus exert a generally unrecognised social power as they identify

3 Erik S. Reinert, ‘The Terrible Simplifiers: Common Origins of Financial Crises and Persistent Poverty in Economic Theory and the New “1948” Movement’, in *Poor Poverty: The Impoverishment of Analysis, Measurement and Policies*, ed. K. S. Jomo and Anis Chowdhury (London: Bloomsbury Academic, 2011).

4 Reinert, ‘The Terrible Simplifiers’.

5 Reinert, ‘The Terrible Simplifiers’, 13.

6 Manuel Castells, ‘A Network Theory of Power’, *International Journal of Communications* 5 (2011): 773–87; Manuel Castells, *The Rise of the Network Society: The Information Age: Economy, Society and Culture*, vol. I (Malden, MA: Blackwell, 1996); D. S. Grewal, *Network Power: The Social Dynamics of Globalization* (New Haven, CT: Yale University Press, 2008).

data and categories for exclusion and employ techniques for variability reduction. The largely unconscious result is to insert their world view in a quantified form, which can obscure—or deny—the existence of alternatives. With each increase in database complexity, the embedded social power is enhanced—as well as being further hidden. Yet mathematical manipulations are usually considered to be value-free.

In the late twentieth century, the quantification of things, processes, activities and events was required for evaluation of, and decisions about, programs and policy formulation. This so-called 'data-driven' or 'technical' approach was alleged to be objective and free of ideology. It was embraced by governments, multilateral institutions and businesses as both a cost-cutting measure and a rational response to newly defined commercial and policy concerns at a global level. The increased capacity of computing technologies, combined with major ideological shifts, caused the reallocation of funds within governments and agencies. Interest moved from national and regional foci to analysis of global trends projected into the future. Less favoured and, frequently, prohibited was primary data collection. Soil scientists and those involved with terrain and land use evaluation were no exception, and responded with new approaches to soil classification, interpretation and mapping. They aggregated existing data sets to create databases for subject areas different from the original contexts of data collection. This is a violation of fundamental rules of statistics governing the validity of extrapolation from defined populations. Completely new globalised data sets and databases were also constructed from numbers extracted (mined) or calculated (derived) from measurement-based data sets.

In his analysis of pre-digital European studies of soil erosion rates, John Boardman decried scientific myth-making: quantification ignoring specific soil properties that produced generalised statistics that moved from report to report, such as a European rate of soil erosion.⁷ Chet Bowers characterised digitisation's separation of information from its ecological and cultural contexts as displacing wisdom with data.⁸ However, those who embrace these procedures refer to them as 'knowledge-based' or 'science-based' decision-making.

7 John Boardman, 'An Average Soil Erosion Rate for Europe: Myth or Reality?', *Journal of Soil and Water Conservation* 53, no. 1 (1998): 46–50.

8 C. A. Bowers, *Let Them Eat Data. How Computers Affect Education, Cultural Diversity and the Prospects of Ecological Sustainability* (Athens, GA: University of Georgia Press, 2000), 48.

Application of derived and reconstructed databases led to error in many fields, causing, perhaps most noticeably, significant losses in the commercial sector. 'Bad data' were identified as the culprit.⁹ In response, the subfield of Data and Information Quality emerged.¹⁰ A growing literature critical of the paucity of statistics at a scale relevant to actual people and places questions the accuracy of data informing socioeconomic debate and policy formulation.¹¹ A smaller literature criticises biofuels databases, models and analyses.¹²

Soil data and mapping

Throughout time and around the world people have used soil description and classification to understand and interact with their environment. Locally significant geomorphic features and soil properties defined land use.¹³ Alfred Hartemink et al. provide a general historical overview of formalised soil description in map form.¹⁴ European interest in mapping soils grew in the early twentieth century. The profession of soil surveying developed and, by mid-century, the possibility of constructing a world soil map emerged. As digital technology offered new tools for expressing ideas about soil properties and soil distributions, mathematical models became increasingly central to soil identification and description. The following paragraphs describe the transition from local to global soil classification and mapping.

9 Richard D. De Veaux and David J. Hand, 'How to Lie with Bad Data', *Statistical Science* 20, no. 3 (2005): 231–8, doi.org/10.1214/088342305000000269.

10 C. Batini and M. Scannapieco, *Data Quality: Concepts, Methodologies and Techniques* (Berlin: Springer, 2006); S. E. Madnick, R. Y. Wang, W. L. Yang and H. Zhu, 'Overview and Framework for Data and Information Quality Research', *Journal of Data and Information Quality* 1, no. 1 (June 2009).

11 For example, Center for Global Development, *Delivering on the Data Revolution in Sub-Saharan Africa*, Final Report of the Data for African Development Working Group Center for Global Development (Washington, DC: Center for Global Development, 2014), www.cgdev.org/sites/default/files/CGD14-01%20complete%20for%20web%200710.pdf, accessed 12 March 2016; David Hsu, 'Improving Energy Benchmarking with Self-Reported Data', *Building Research and Information* 42 (2014): 641–56; Mørtten Jerven, *Poor Numbers: How We Are Mislead by African Development Statistics and What to Do About It* (Ithaca, NY: Cornell University Press, 2013); P. Knox Clarke and P. Darcy, *Insufficient Evidence? The Quality and Use of Evidence in Humanitarian Action* (London: ALNAP/ODI, 2014); Reinert, 'The Terrible Simplifiers'.

12 For example, Niclas Scott Bentsen and Claus Felby, 'Biomass for Energy in the European Union—A Review of Bio-energy Resource Assessments', *Biotechnology for Biofuels* 5, no. 25 (2012): 2, www.biotechnologyforbiofuels.com/content/5/1/25, accessed 10 December 2018; Sarah M. Lewis and Maggi Kelly, 'Mapping the Potential for Biofuels Production on Marginal Lands: Differences in Definitions, Data and Models Across Scales', *ISPRS International Journal of Geo-Information* 3 (2014).

13 Benno P. Warkentin, ed., *Footprints in the Soil: People and Ideas in Soil History* (Amsterdam: Elsevier, 2006); Kate B. Showers, 'A History of African Soils: Perceptions, Use and Abuse', in *Soils and Societies: Perspectives from Environmental History*, ed. J. R. McNeill and V. Winiwarter (Isle of Harris: White Horse Press, 2006). Ethnopedology Special Issue, *Geoderma* 11 (2003); Antoinette M. G. A. Winklerprins, 'Local Soil Knowledge: A Tool for Sustainable Land Management', *Society and Natural Resources* 11, no. 7 (1999): 151–66; P. Sillitoe, 'Knowing the Land: Soil and Land Resource Evaluation and Indigenous Knowledge', *Soil Use and Management* 14 (1998): 188–93, doi.org/10.1111/j.1475-2743.1998.tb00148.x.

14 Alfred E. Hartemink, Pavel Krasilnikov and J. G. Bockheim, 'Soil Maps of the World', *Geoderma* 207–8 (2013).

World soil data and map

In response to a 1960 International Soil Science Society (ISSS) recommendation to 'harmonize and synthesize' world soil knowledge,¹⁵ field staff of the United Nations (UN) Food and Agriculture Organization (FAO) conducted soil surveys to fill gaps within and between existing soil maps. The resulting Soil Map of the World, published by the FAO in collaboration with the United Nations Social, Educational and Cultural Organization (UNESCO), was based on direct information linked to mapping units defined by soil types.¹⁶ It was produced as a series of map sheets and explanatory volumes between 1968 and 1978.

Although most European scientific soil survey information was incorporated, large areas of the world had not been assessed. Indigenous soil classification systems were not drawn upon because they were unknown, too local in reference, or unrecognised. Assumptions were made about the properties of so-called tropical and Mediterranean soils and climates.¹⁷ For example, the 1977 Map Sheet 7, covering the continent of Africa (22 per cent of the earth's surface), contains a qualifying legend: 7 per cent of the diverse continent had been covered by coarse- or medium-scale survey maps 'with some grounds for accuracy'; 38 per cent by reconnaissance maps showing soils in relation to other features such as climate, geomorphology or vegetation; and 55 per cent of the continent was 'virtually unknown'.¹⁸

Despite criticism for misinformation and omissions,¹⁹ all subsequent world soil maps have been based upon these FAO/UNESCO map sheets.²⁰ As electronic technology replaced boots and shovels, field survey maps and laboratory analyses were labelled 'legacy data', and the Soil Map of the World was referred to as 'the paper map'. However, there was recognition that the richness of the paper map's supporting data had been hidden by presentation and needed revision. Expansion of geographical information system (GIS) technologies led to digitisation of the 'often incomplete and not fully elaborated' world soil data.²¹ This began with the Environmental Systems Research Institute in Vector format, and led to the 1991

15 Hartemink et al., 'Soil Maps of the World', 261.

16 Food and Agriculture Organization/United Nations Educational, Scientific and Cultural Organization, *Soil Map of the World*, vol. VI: *Africa* (Paris: UNESCO, 1977).

17 R. Lal and P. A. Sanchez, *Myths and Science of Soils of the Tropics*, Soil Science Society of America Special Publication, 29 (Madison, WI: Soil Science Society of America, 1992); A. T. Grove, 'The African Environment, Understood and Misunderstood', in *The British Intellectual Engagement with Africa in the Twentieth Century*, ed. Douglas Rimmer (New York: St Martin's Press, 2000).

18 FAO/UNESCO, *Soil Map of the World*.

19 Bouman, quoted in F. O. Nachtergaele, *From the Soil Map of the World to the Digital Global Soil and Terrain Database: 1960–2002* (Rome: FAO, 2004), www.css.cornell.edu/faculty/dgr2/Docs/WRB/SoilMapWorld.pdf, accessed 15 June 2015; Hari Eswaran, Fred Beinroth, J. Kimble and T. Cook, 'Soil Diversity in the Tropics: Implications for Agricultural Development', in *Myths and Science of Soil of the Tropics*, ed. R. Lal and Pedro A. Sánchez (Madison, WI: Soil Science Society of America, 1992); A. T. Grove, 'The African Environment, Understood and Misunderstood'; A. T. Grove and O. Rackham, *The Nature of Mediterranean Europe* (New Haven, CT: Yale University Press, 2001).

20 Nachtergaele, *From the Soil Map of the World*.

21 *ibid.*

FAO Digitized Soil Map of the World (DSMW).²² Focusing exclusively on soil data, the DSMW allowed analysis by individual country, but required an ARC-Info GIS system for full utilisation. Faster versions followed in the 1990s. The inverse relationship between fieldwork funding and information technology developments continued. FAO's soil scientists embraced technological advances in remote sensing, computers, terrain analysis, geostatistics, GIS, data integration and instrumentation to achieve 'unprecedented reliability and utility in digital soil maps'.²³

Globalised soils

The neoliberal economic model came to dominate late twentieth-century planning and policy regimes. Trade was conceptualised in terms of global markets to be supplied by regions most suited to each commodity's production. Places with constraints on production were devalued, since alternative locations could be found in other regions or on other continents better suited to meet the scale of production required to satisfy global demand. Data and statistics collected to support strengthening national self-sufficiency and national economic agendas were inappropriate for analysis at a global scale because of the heterogeneity reflected in their detail. Changing scales mandated data modification, which was made possible by expanding information technology.

Global soil, terrain and landscape analysis begins with data aggregated to a national scale. Member nations of the UN produce data characterising their populations, landscapes and economic activities, which various UN agencies are mandated to acquire, curate and make available to the public. Both what to collect and methodologies of collection have varied from place to place, and over time. National data have, therefore, been of uneven quality.

UN statisticians published not only qualifying paragraphs with their databases, but also reports analysing their limits, with cautions about (ab)use. For example, in 2007, the UN Statistics Division (UNSD) surveyed the 192 member nations about their environmental statistics. Less than 50 per cent of non-European and non-North American countries had any information that could be considered to be environmental. Of the eight reasons listed, availability of data was the most common impediment to data collection.²⁴

The constraints on fieldwork and complications of inadequate national statistics on global database construction were circumvented by applying digital techniques and sophisticated mathematics. Database construction became database

22 *ibid.*

23 Food and Agriculture Organization, FAO Soils Portal, 'Legacy Soil Maps and Soils Databases' (2019), www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/en/.

24 United Nations Statistics Division, *Global Assessment of Environment Statistics and Environment—Economic Accounting* (New York: United Nations Department of Economic and Social Affairs, Statistics Division, 2007), unstats.un.org/unsd/statcom/doc07/Analysis_SC.pdf, accessed 16 October 2015.

creation, and map-making moved from largely pedantic exercises stimulated by exploration and discovery to a more descriptive activity better suited to the needs of program-level evaluation and analyses by non-soil scientists. To assist users from diverse backgrounds and interests, soil scientists created digitised databases and maps, as well as websites with soils portals emphasising ease of access.

Pedometrics: A digital soil mapping tool

Digital soil mapping's primary approach was the mathematical process of deriving data, called pedometrics. The concept was developed in the late 1970s when some FAO soil scientists, using a land evaluation approach, devised a set of rules for the expansion of individual, direct observations to encompass wider areas. Pedometrics consisted of two major functions: pedotransfer and taxotransfer.²⁵ Pedotransfer is 'the mathematical relationship between two or more soil parameters which shows a reasonable high level of statistical confidence. [It is used to] facilitate estimation of a non-measured soil parameter from one or more measured ones'.²⁶

Common relationships identified included those between pH and base saturation; cation exchange capacity and clays and organic matter; and salinity, pH and exchangeable sodium percentage.²⁷ Pedotransfer functions were to be used only when soil profile data were available. In contrast, taxotransfer is an act of simplification:

[the] estimation of soil parameters based on modal soil characteristics of soil units, as derived from a combination of their classification name or taxon (which, by definition often implies a certain range for a number of properties). This procedure relies upon expert knowledge, empirical rules, and a statistical analysis of a large number of soil profiles belonging to the same taxon.²⁸

Taxotransfer procedures were first used by FAO in 1992.²⁹

In the 1980s a working group of the International Union of Soil Sciences developed a common language for naming soils: the World Reference Base for Soils. Its purpose was:

to serve as a framework through which ongoing soil classification throughout the world can be harmonized [to] reach international agreement on the major soil groups to be recognized at a global scale as well as on the criteria and methodology to be applied for defining and separating them.³⁰

25 F. O. Nachtergaele, 1976, quoted in Nachtergaele, *From the Soil Map of the World*.

26 Nachtergaele, *From the Soil Map of the World*, 6.

27 *ibid.*, 7.

28 *ibid.*

29 *ibid.*

30 J. A. Deckers, P. M. Driesssen, F. O. Nachtergaele and O. C. Spaargaren, *World Reference Base for Soils in a Nutshell 2001*, Research Report, 7 ([Ispra]: European Soil Bureau Network, 2002), www.researchgate.net/publication/267642515_World_Reference_Base_for_Soil_Resources_-_in_a_nutshell/download, accessed 17 May 2017.

These acts of simplification that created coarse-grained global data involved deliberate reductions of information. The goal was to ‘make soil survey, classification, and land evaluation as objective as possible’.³¹

The development of pedometrics illustrates not only digital approaches, but also the normality of embedding assumptions and opinions in global soil, land and terrain databases by mixing measurement data (from legacy databases and other sources) together with synthesised values. Gaps in globalised data sets were filled with subjective/socially constructed, rather than measurement-based, numbers. These are difficult for end users to recognise, and thus have become fundamental components of the coarse-grained global analyses.

The potential for applying pedometrics seemed limitless. Subjectivity categorised as ‘expert knowledge’ and ‘empirical rules’ could be quantified and interwoven with measurements to build virtual soil profiles serving as proxies and estimates in databases and maps at different scales. These synthetic soil databases could be digitised and combined with digitised remote sensing and soil survey data to provide more complex and comprehensive databases and analyses. However, digitisation embedded a contradiction in soil maps. The technology requires discrete and static categories, which violates the fundamental concept of soil bodies as a continuum constantly forming and changing in relationship to biogeophysical forces. Not only was site-specific detail eliminated to achieve ‘harmonisation’, but sites themselves were eliminated—and replaced by grid cells.

Soil and TERrain (SOTER) database

The first application of new soil data construction methods to compile a global soil data set was the scale-independent SOil and TERrain (SOTER) digital database.³² Launched by ISSS in 1986, the SOTER Digital Database Project was implemented by a collaborative group from the UN Environment Programme (UNEP), FAO and the International Soil Resource Information Centre (ISRIC) to create, ‘at a 1:500,000 scale a World Soils and Terrain Database with digitised map units whose attribute data was in a standardised format’.³³

Explicitly prohibited from making new field soil surveys, the group’s activities were to be based upon existing information classified under a new, universal map legend. Consistent pedotransfer rules derived from the ISRIC’s World Inventory

31 FAO, *SOTER-Global Soil and Terrain Database* (Rome: FAO Land Resources Division, 2015), www.fao.org/nr/land/databasesinformation-systems/soter/en, accessed 16 October 2016.

32 Vincent van Engelen and Alfred Hartemink, ‘The global SOils and TERrain database (SOTER)’, *ACLEP Newsletter* 9, no. 1 (April 2000): 22–26, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.1264&crep=rep1&type=pdf, accessed 10 December 2018.

33 *Soil and TERrain Database (SOTER) Programme*, www.isric.org/projects/soil-and-terrain-soter-database-programme, accessed 27 January 2016.

of Soil Emission Potentials (WISE) soil profile database filled gaps in existing measured soil profile data.³⁴ Each SOTER unit represents a unique combination of terrain and soil characteristics. The units are described by 118 attributes, which are compatible with data stored in the FAO/ISRIC soil database.³⁵ The resulting database consisted of sets of files for use in a Relational Database Management System and in GIS.³⁶ Initially designed for use at a scale of 1:1 million, it has been applied at larger scales.³⁷

The SOTER program produced a series of 'digital products'³⁸—including a soil map of Africa. Applying pedotransfer rules, soil scientists constructed secondary data sets (SOTWIS). These had a wide range of applications, including layers in the FAO-led Harmonized World Soil Database.³⁹ One of the original goals of replacing the soil survey-based 1978 paper map with a Global SOTER was to increase world soils' knowledge. Unfortunately, this has not been achieved because SOTER is a mixture of measurement and derived data. For example, SOTER Central Africa covers the 2.4 million km land area of Burundi, the Democratic Republic of Congo and Rwanda using 244 SOTER units, corresponding to 504 polygons. Each SOTER mapping unit was 'characterized using 167 real soil profiles selected by soil experts as being regionally representative of these [SOTER] units [along with] 129 virtual profiles for which only the FAO classification was known'.⁴⁰

There are no field measurement data for these profiles. Data gaps were filled using taxotransfers.⁴¹

Users of SOTER and its 'products' should remember that, first, SOTER information is synthetic rather than measurement-based; second, SOTER units are not linked directly to individual soil bodies or specific places in landscapes; third, SOTER units were meant for use at coarse scales; and fourth, SOTER products were largely designed for use by policymakers and others interested in broad general statements without accuracy at local levels.

34 *SOil and TERRain Database (SOTER) Programme.*

35 Van Engelen and Hartemink, 'SOTER'.

36 *SOil and TERRain Database (SOTER) Programme.*

37 *The global SOils and TERRain database (SOTER).*

38 International Soil Resource Information Centre, 'From SOTER to SOTWIS', Soil and Terrain (SOTER) Database Programme, www.isric.org/projects/soil-and-terrain-soter-database-programme, accessed 20 May 2017.

39 *SOil and TERRain Database (SOTER) Programme.*

40 N. H. Batjes, *SOTER-based soil parameter estimates for Central Africa—DR of Congo, Burundi and Rwanda*, (SOTWIScaf ver 1.0) Report 2007/02 (Wageningen: ISRIC—World Soil Information, 2007), iii, www.isric.org/sites/default/files/isric_report_2007_02.pdf, accessed 17 May 2017.

41 Batjes, *SOTER-based Soil Parameters.*

Worldwide agricultural databases

The shift in interest from local to global scales was also reflected in agricultural databases. Digital technologies provided the tools not only for change in scale, but also for broader applications. In the late twentieth century, collaborating economists, soil scientists, ecologists and modellers applied these techniques to convert worldwide national and regional agricultural databases collected in response to concerns about poverty, hunger and food insecurity. They were transformed into databases of globalised statistics to address different problems formulated at a global scale. The following sections illustrate these processes.

World Agriculture: Towards ... series

The FAO's *World Agriculture: Towards...* series provides forecasts of future crop, feed and fibre production. The earliest publication (1970) relied on expert opinion as the primary data source. It included tabulated 'optimistic estimates' of agricultural production in 'developing countries' to make assessments at regional levels.⁴² In contrast, the 1988 *World Agriculture: Towards 2000* used national data in the economic quantitative framework of Supply Utilization Accounts to build scenarios. The result was 'more moderate' quantified rates of agricultural growth consistent with national goals of attaining self-sufficiency.⁴³ Unlike most contemporary global studies, 'each global statement is derived from a summation of, and can be decomposed back into, a number of constituent single-country or commodity statements'.⁴⁴

Systematic application of what Reinert labelled as abstraction⁴⁵ provided reproducible assessments. Although still including it, expert opinion was used in conjunction with systematic data produced by a model employing reported measurement statistics. In the late 1980s, when information technology innovations were in their infancy, the choice was between either:

- a roughly estimated formal model with much less commodity input and country detail and
- a huge model with all the detail of this study but with the bulk of the parameters and coefficients being 'guestimates' rather than data.⁴⁶

Beginning with the *Towards 2015/30* studies in 2000, the series editors employed GIS technologies and the global databases described in the following sections.⁴⁷

42 Nikos Alexandratos, ed., *World Agriculture: Towards 2000, An FAO Study* (Rome: FAO; New York: New York University Press, 1988), 2.

43 Alexandratos, *Towards 2000*, 2.

44 *ibid.*, 277.

45 Reinert, 'The Terrible Simplifiers'.

46 Alexandratos, *Towards 2000*, 278.

47 FAO, *World Agriculture: Towards 2015/30: Interim Report* (Rome: FAO Global Perspectives Unit, 2000).

Global Assessment of Soil Degradation (GLASOD)

The GLASOD grew from ISRIC soil scientists' fears that inadequate attention was being paid to observed, but unquantified (and largely undefined), soil degradation. UNEP officers agreed that a quick (rather than scientific) document should be prepared for the 1992 UN Conference on Environment and Development (UNCED) to raise awareness among policy designers and decision-makers about soil damage caused by improper land use.⁴⁸ With UNEP funding in 1987, ISRIC coordinated a three-year project using an 'expert system approach' to produce a Mercator projection-based world geographical map that would provide 'an easily displayed simple visual expression'.⁴⁹

First, the ISRIC soil scientists divided the world into 21 regions, then identified institutions and individual soil and environmental scientists considered to have expert knowledge. After selecting regional coordinators, the project administrator sent each participant a structured survey questionnaire, simplified geographic base map and guidelines for 1) describing the type and degree of degradation and 2) estimating the average degradation rates in the previous five to 10 years. The lead ISRIC soil scientist defined soil degradation as 'a process that describes human-induced phenomena which lower the current and/or future capacity of the soil to support human life'.⁵⁰ Respondents were asked to indicate regions where 'the balance between the attacking natural forces of climate and the natural resistance of the terrain against these forces has been broken by human intervention'.⁵¹

The 290 country collaborators working at a scale of 1:7,500,000 submitted results with an average scale of 1:10 million. Cartographers produced the final 1:10 million map, which was published in 1990. When digitised in 1991, the map was released with an attribute database and supplementary statistics.⁵² However, the small working scale produced such a coarse-grained analysis that 'soil degradation of local importance cannot be delineated'; when combined into a single map, entire mapping units disappeared.⁵³ Responding to criticism about difficulties in determining the amount of land under threat, a 1991 second revision of the 1990 Explanatory Note was published with an annex that 'discusse[d] the actual areal extent'.⁵⁴

48 L. R. Oldeman, ed., *Guidelines for General Assessment of the Status of Human-Induced Soil Degradation*, Working Paper and Preprint 88/4 (Wageningen: International Soil Reference and Information Centre, 1988); L. R. Oldeman, R. T. A. Hakkeling and W. G. Sombroek, *World Map of the Status of Human-Induced Soil Degradation: Explanatory Note* (Wageningen: International Soil Reference and Information Centre; Nairobi: UNEP, 1990); L. R. Oldeman, R. T. A. Hakkeling and W. G. Sombroek, *World Map of the Status of Human-induced Soil Degradation: An Explanatory Note*, 2nd ed. (Wageningen: International Soil Reference and Information Centre; Nairobi: UNEP, 1991; UNSD, 2004).

49 Oldeman et al., *World Map*.

50 Oldeman, *Guidelines*, 2.

51 Oldeman et al., *World Map*, 7.

52 United Nations Statistics Division, *Land Degradation and Land Use/Cover Data Sources*. Working Document, 31 December 2004 (New York: United Nations Department of Economic and Social Affairs, Statistics Division, 2004).

53 Oldeman et al., *Guidelines*, 2.

54 Oldeman et al., *World Map*, vi.

The GLASOD map and database became easily accessible after digitisation and incorporation into UNEP's Global Resource Information Database (GRID) for assessing desertification. The GLASOD map appears in many publications and the database has been widely cited and used. For example, a chapter on global soil degradation in UNEP's *1992 State of the Environment Report* was based on the GLASOD database. Despite its authoritative appearance and institutional publication, both the initial collection of information and its synthesis into a global map involved quantification of subjective assessments and processes of simplification with deliberate elimination of information.

Global models and model-created databases

The emerging digital technologies did not only enable the expansion and proliferation of globalised databases. They also stimulated analysis of soil capacities and capabilities at a global scale through modelling. Because models are mathematical relationships, detailed knowledge of the soils and their properties at specific places is not required. Instead, modellers and their models operate in the mathematically constructed, abstract world of pixels and grid cells.

Models are developed by individuals or groups to analyse relationships between or among carefully defined entities. The scope of the model must be agreed upon, followed by identification of the best parameters with which to characterise its components, and how they relate to one another. Once this intellectual framework has been determined, mathematical expressions and relationships can be devised to make a quantitative analytical framework. The resulting basic model can be modified subsequently for different applications, or in response to changing data availability. Models derived from the same base (standard) constitute a family of models.

Careful definition is required before quantification can begin. No model can include everything, but the deliberate or inadvertent exclusion of crucial factor(s) can cause models to produce analyses that, although mathematically correct, are imprecise—or completely wrong. The decision-making process is inherently biased, since it depends upon each modeller's cultural and philosophical perspectives and values (e.g. academic discipline, social concerns, economic beliefs). These will be consciously or unconsciously expressed as a problem is framed and relationships described. Because modelling involves continuous choices between inclusion and exclusion in both construction and application, it is also an exercise in social power. No matter how mathematically complex, a model cannot be value-free or objective.

Agro-ecological zone methodology and databases

The agro-ecological zone (AEZ) methodology is based on principles of land evaluation. It was designed as a tool for determining agricultural production potentials and the carrying capacity of the world's land areas. After its concepts and techniques—including pedometrics—had been field tested in regional studies, this methodology became FAO soils field staff and associates' primary approach to land resource assessment.⁵⁵ The AEZ framework was digitised when the 1995 DSMW became available. Digitisation created a new category of land defined both by and in technology: agro-ecological zones. An AEZ consists of all parts of grid cells on a georeferenced map having uniform soil and climate characteristics. The digitised AEZ methodology begins with the evaluation of an AEZ in terms of a list of 21 food, fibre and fodder crops and pasture grasses under rain-fed conditions using three specified levels of crop production technology.

The advent of GIS and increased computing capabilities underlay a collaboration between the International Institute for Applied Systems Analysis (IIASA) and the FAO. The purpose was to operationalise the AEZ methodology by building a GIS-based modelling framework that would, according to its documentation:

- provide a standardized framework for characterization of climate, soil and terrain conditions relevant to agricultural conditions;
- identify crop-specific limitations of prevailing climate, soil, and terrain resources under assumed levels of inputs and management conditions;
- provide the framework for various applications, such as quantification of land productivity, extents of land with rain[-]fed or irrigated cultivation potential, estimation of the land's population supporting capacity, and multi-criteria optimization of land resources use and development.⁵⁶

The collaborators created original databases of derived statistics using spatial biophysical and socioeconomic data sets for use in AEZ analysis.⁵⁷

The model's creators claimed that, operating at the grid cell rather than local landscape level, the AEZ model was capable of 1) simulating land resources availability; 2) assessing farm and overall management options; 3) estimating crop production potentials under different cropping patterns and land utilisation types

55 International Institute of Applied Systems Analysis, *Global Agro-Ecological Assessment for Agriculture in the 21st Century*, IIASA, webarchive.iiasa.ac.at/Research/LUC/SAEZ/index.html, accessed 7 April 2016; IIASA/FAO, *Global Agro-Ecological Zones (GAEZ v 3.0)* (Laxenburg: IIASA; Rome: FAO, 2012); Nachtergaele, *From the Soil Map of the World*.

56 IIASA, *Agro-Ecological Zones. Methodology. Global Agro-Ecological Zones*, IIASA, www.iiasa.ac.at/Research/LUC/GAEZ/method/method.htm, accessed 27 September 2007.

57 N. H. Batjes et al., *Soil data derived from WISE for use in global and regional AEZ studies (version 1.0)*, 1997, pure.iiasa.ac.at/id/eprint/5266/1/IR-97-025.pdf, accessed 27 January 2016.

(LUTs); and 4) estimating potential land availability.⁵⁸ These simulations and estimates are defined by the model's parameters (e.g. crop, soil, climate attributes), assumptions (e.g. land use value, management possibilities) and the quality and scale of the databases employed.

The model allowed the distribution of computations at fine-gridded intervals over the entire globe, providing a 'standardized framework for the characterization of climate, soil and terrain conditions relevant to agricultural production'.⁵⁹ A geo-referenced AEZ database,⁶⁰ created at the FAO in Rome, consisted of 2.2 million grid cells that included:

- [an] environmental conditions component comprised of georeferenced global databases for climate, soils, terrain and elevation, and
- [a] spatial land use and land cover component covering forests, protected areas, irrigated areas, population distribution and density, land required for habitation and infrastructure, estimates of cropland, grazing land and sparsely vegetated or barren land, and farming systems.⁶¹

Initial AEZ grid-cell analysis indicated that, 'at a global scale', there were enough 'biophysical resources' to produce the required amounts of food, feed and fibre for future generations. However, the documentation noted 'profound concern' about 'the capacity in several regions and countries with limited land and water resources'.⁶² The AEZ model was also used to generate new data sets for further analysis using mathematical manipulations of largely derived data.⁶³ When published in 2000, AEZ became the base for an IIASA/FAO family of models.

Global agro-ecological zones model

In 2000, emerging global databases enabled the extension of AEZ to global agro-ecological zones (GAEZ) analysis. GAEZ assessments are also estimated by grid cell and aggregated to national, regional and global levels. The original AEZ database was expanded to include 'interactive multi-criteria model analysis' to make AEZ more accessible to decision-makers for 'rational land use planning'.⁶⁴ A database of 'viable landuse options' was constructed from 'limited' socioeconomic

58 IIASA/FAO, *Global Agro-Ecological Zones*.

59 IIASA, *Global AEZ-2000 home*, IIASAnalysis, www.iiasa.ac.at/Research/LUC/GAEZ/faq.htm, accessed 27 September 2007; IIASA/FAO, *Global Agro-Ecological Zones*.

60 Nikos Alexandratos, ed., *World Agriculture: Towards 2010, An FAO Study* (Rome: FAO; Chichester: John Wiley and Sons, 1995), 153.

61 *Land Resources Agro-ecological Zoning System*, Release 2002, www.fao.org/nr/land/databasesinformation-systems/aez-agro-ecological-zoning-system/en, accessed 23 January 2016.

62 Günther Fischer et al., 'Agro-ecological Zone Assessment', in *Land Use Land Cover and Soil Sciences*, vol. III, ed. W. H. Verheye (Oxford: EOLSS Publications, 2006), 5.

63 IIASA/FAO, *Global Agro-Ecological Zones*.

64 IIASA, *GAEZ*, IIASA, 2014, www.iiasa.ac.at/web/home/research/modelsData/GAEZ/GAEZ.en.html, accessed 13 October 2015.

information—certain land use practices identified as ‘land utilization types’.⁶⁵ Availability of new data sets and databases enabled new versions of GAEZ. Each time it was updated, the problems addressed, the size of the databases and the number of results increased. GAEZ databases became the ‘agronomic backbone’ for a variety of applications, including the quantification of land productivity.⁶⁶

Version 3.0, publicly released in May 2012, marked a significant departure from, rather than a simple extension or modification of, the original GAEZ model. Along with revised and updated components, there was a new procedure for ‘spatially downscaling agricultural production statistics applied to produce a global gridded inventory of year 2000 agricultural yields and production’.⁶⁷

There were also new statistics and databases, including observed climate data and climate scenarios, derived slope and aspect data, and the new spatially Harmonized World Soil Database.⁶⁸ Despite the use of the latest available global data sets and databases, either in their entirety or as sources of data for the newly derived data sets designed for the model:

the quality and reliability of these datasets are known to be uneven across regions. Reliability of the information contained in the database is variable ... the agronomic data ... contain generalizations necessary for global applications ... in particular, assumptions on occurrence and severity of some agro-climate related constraints to crop production would undoubtedly benefit from additional verification and data.⁶⁹

Version 3.0 left the original AEZ mandate of national crop production policy formulation and moved to the level of globalised analysis. In particular, it could address issues associated with global change processes. The required global estimations had been impossible to construct using conventional statistics. While methods to approach spatially bounded problems

are based on the ability to obtain observations from unknown true probability distributions ... the new problems require recovering information from only partially observable or even unobservable variables[.]⁷⁰

65 IIASA, *GAEZ*.

66 IIASA/FAO, *Global Agro-Ecological Zones*, 2.

67 IIASA/FAO, *Global Agro-Ecological Zones*, xii.

68 IIASA/FAO, *Global Agro-Ecological Zones*.

69 FAO, ‘What are the GAEZ limitations?’ *Frequently Asked Questions (FAQ)*. GAEZ: Global Agro-Ecological Zones, FAO, www.fao.org/nr/gaez/faqs/en, accessed 23 January 2016.

70 IIASA/FAO, *Global Agro-Ecological Zones*, 1.

This necessitated the development of novel

sequential balancing procedures ... that rely on appropriate optimization principles ... e.g. cross-entropy maximization, and combine the available samples of real observations in the locations with other ‘prior’ hard (statistics, accounting identities) and soft (expert opinion, scenarios) data.⁷¹

The results of GAEZ assessments are presented as authoritative-looking maps and tables at national, regional and global levels. They resemble those based on conventional field measurement. Yet the databases employed have been constructed from mathematical manipulations that approach Reinert’s state of excessive simplification⁷²—practical irrelevance. Nevertheless, GAEZ is promoted as a tool for land use planning and has been endorsed by the International Panel on Climate Change (IPCC) as a tool for climate change research.⁷³

Land resource potential and constraints

Land Resource Potential and Constraints at Regional and Country Levels was the first application of preliminary results from GAEZ analysis. It was an early attempt to apply ‘electronic data-processing through geographical information systems (GIS) and modelling’ for analysis of existing data on land potential, soils and landforms in terms of climatic parameters and integrated with population and population density data.⁷⁴ The result was a series of tables not supported by GIS. The analysts wrote an explanatory text that, together with the tables, was published as *World Resources Report 90*, and posted on the FAO website.⁷⁵

The report was organised around five themes: soil constraints, drylands and steplands, land degradation, potential arable land for rain-fed agriculture, and country ranking according to land resource potential and constraints. These were expressed in tables, bar graphs and maps in the text and appendices. Each section included discussion of the methodology and limitations of the data. According to the authors, the report was distinctive among FAO publications

71 IIASA/FAO, *Global Agro-Ecological Zones*, 1.

72 Reinert, ‘The Terrible Simplifiers’.

73 *United Nations Framework Convention on Climate Change, Compendium on Methods and Tools to Evaluate Impacts of, and Vulnerability and Adaptation to, Climate Change*. Final Draft Report. SC 10341, (Bonn: United Nations Framework Convention on Climate Change Secretariat, 2005).

74 FAO, *Land Resource Potential and Constraints at Regional and Country Levels* (Rome: FAO, 2000), 1; Nachtergaele et al., *Terrastat: Global Land Resource GIS Models and Databases for Poverty and Food Insecurity Mapping*, 2, beta version 1.0, 2002.

75 A. J. Bot, F. O. Nachtergaele and A. Young, *Land Resource Potential and Constraints at Regional and Country Levels* (World Soil Resources Reports, 90) (Rome: FAO Land and Water Development Division, 2000, www.fao.org/3/a-x7126e.pdf, accessed 11 December 2018).

because it was worldwide in scope (covering 160 nations), enabling comparison 'between developing countries, Commonwealth of Independent States and Baltic States and developed countries'.⁷⁶

Land Resource Potential and Constraints introduced, defined and quantified technology-derived terms for land that have become part of global analyses. 'Potential arable land' describes land not used for agriculture but, according to model parameters, is suitable or marginal for the rain-fed cultivation of any one of 21 reference crops.

Assessments of potential arable land have been central to many agricultural and biofuels analyses, despite the text's caveat that

the overall validity of the methods employed, and hence the results obtained, have been questioned ... the results presented ... should therefore be read in conjunction with the reservations made in the section 'The reliability of estimates of available cultivable land'.⁷⁷

The report also included numerical estimates of 'gross potential arable land' and 'net potential land'. *Land Resource Potential* is the source of the assertion that sub-Saharan Africa and South and Central America contain most of the increase in global potential arable land.⁷⁸ Subsequent discussion mentions that since forest cover is important, and particularly rainforest, the estimated acreage should be reduced. Both the Amazon and Congo basin rainforests had been included in the calculated potential arable land.

The report's authors also consciously used the emerging digital technologies to assess and confirm biases and expert opinion:

in addition to the analysis of specific kinds of data, some calculations are made between physical resource constraints, and degradation, and population, to explore the extent to which intuitive relationships are confirmed.⁷⁹

Subjectivity and estimates are at the core of *Land Resource Potential*, making all amounts and conclusions extremely unreliable. The analysis relied on the previously discussed GLASOD database, derived from expert opinion, FAO estimates of available land calculated from the coarse scale non-fieldwork based DSMW, land-use estimates prepared by FAO staff for *World Agriculture: Towards 2010*,⁸⁰ and the unverified preliminary results of the GAEZ model. This report, therefore, authoritatively presents completely manufactured numbers that have been accepted

76 FAO, *Land Resource Potential*, 1.

77 *ibid.*, 37.

78 *ibid.*, 38.

79 *ibid.*, 1.

80 Alexandratos, *Towards 2010*.

as fact. *Land Resource Potential and Constraints* has been widely cited. The concept of potential arable land is frequently used in academic and policy research, and values from the report's tables have proliferated in scientific and socioeconomic biofuels research and policy literature.

TERRASTAT global land resources GIS models and databases

Among the responses to the 1996 FAO World Food Summit was the TERRASTAT database. FAO soil scientists created it as part of a Poverty and Food Insecurity Mapping project funded by the government of Norway. TERRASTAT's intended use was to analyse, at global scales, where people live in relation to agro-ecological zones, marginal or productive lands, and major food production systems—particularly those important to food security for the rural poor.⁸¹

The TERRASTAT database contains country statistics of soil, terrain, climatic and agricultural land use limitations and potential extent linked in tabular format with population and population density in each of the identified biophysical zones. Released in 2001 by the FAO, it was published the following year in a user-friendly two CD-ROM set containing computer programs that displayed and exported different biophysical themes linked to population density. Other applications of the databases and models were subsequently linked to the Land Degradation Assessment in Drylands project (LADA).⁸²

TERRASTAT was based on the *Land Resource Potential and Constraints* study. Its creators used the first published version of GAEZ, which included more recent biogeophysical databases and a newer global population database. The smallest unit of analysis was an agro-ecological cell, 'a unique lgp [length of growing period]-thermal-climate-soil type-texture class-soil phase and slope combination'.⁸³ TERRASTAT includes tables at national and regional levels related to six themes: major soil constraints, soils in deserts and dryland areas, population distribution, steplands analysis, land degradation severity, and human-induced land degradation due to agricultural activities. Tables further elaborate the concept of 'potential arable land'.

Despite updated databases and model components, TERRASTAT shared with its predecessor both intellectual approaches and problems of data quality and reliability. It also has had widespread application. By posting tables and text separately on the FAO website, TERRASTAT's developers enabled the downloading of tables without explanatory or qualifying notes. In this way, estimates, quantified opinions and misleading calculations garnered the appearance of fact supported by the

81 *Terrastat*, 1.

82 *ibid.*, 2.

83 *ibid.*, 31.

institutional authority of FAO. The tables were influential in discussions of land availability for increased crop production in general, and the possibility of agrofuels and biofuels cultivation in particular.

Scholarly critique ensued. In late 2010, qualifying text had been added to the web page above the list of downloadable tables; by late 2015, the TERRASTAT database was less easily accessible on the FAO website. Nevertheless, the term—and belief in the existence of—potential arable land persisted. It has been adopted by analysts and researchers, and proliferated throughout the agriculture and land use literature.

Globalised land resource assessment and GIS-based biofuels production

Critics asserted that biofuels cultivation exacerbated global land shortage. In response, advocates reinterpreted data supporting earlier analyses of agricultural land bases, and employed GIS-based models to search for vacant land. The amount of available land identified depended not only upon the definition of land and criteria applied to identify land use, land rights and land ownership, but also on climate, terrain and soil characteristics. The earliest estimates came from simple analyses of available databases—first at the national and regional levels, then at the global scale. Sarah Lewis and Maggi Kelly's review of GIS-based biofuels analyses found that despite general theoretical agreement that marginal land was land unsuited for agriculture, each set of authors had a different working definition, which determined choices made about and within models.⁸⁴ In addition, results from (the majority of) models using Boolean overlays differed depending upon the order in which inputs had been made. Only two studies used fuzzy logic set theory, one being GAEZ. Globalised GIS-based analyses have bolstered the biofuels industry's claim to being a constructive response to the climate change crisis. However, Lewis and Kelly emphasised the importance of connecting theoretical calculations with 'real world conditions' of the balance between fuel, food and conservation.⁸⁵

Consumers of global biofuels analyses should recognise that biofuels feedstocks cannot be cultivated on grid cells. In reality, the crops proposed for fuel production range from oil seeds to woody plants that grow in drylands, wetlands, temperate zones or the tropics. Each species has specific growth requirements, and each landscape offers clearly defined biogeophysical and sociocultural conditions. Environmental historians and geographers have documented the unintended—and often disastrous—consequences of ignorant or ill-informed interventions in

84 Sarah M. Lewis and Maggi Kelly, 'Mapping the potential for biofuels production on marginal lands: differences in definitions, data and models across scales', *ISPRS International Journal of Geo-Information* 3 (2014): 430–59.

85 Lewis and Kelly, 'Mapping'.

landscapes and the societies they support.⁸⁶ Special attention has been paid to the large-scale economic and development projects of high colonialism on the African continent, but other peoples and empires in other times and on other continents have left their sociocultural and environmental marks as well. Discussing the fallacy of the notion of pristine nature and environments untouched by human existence, and the contrasting reality of humans, other life forms and geophysical forces as co-creators of landscapes and their ecosystems, is far beyond the scope of this paper. Taken together, the large body of research about particular places and times attests to the critical importance of the local, and the need to appreciate the power of heterogeneity. While ecological theory recognises both diversity as central to creating stability and resilience, and complexity as an inherent quality of environmental interactions, modellers consider these to be unnecessarily complicating detail and seek to exclude them from consideration.

Summary and conclusions

The primary response to recognition of human implication in changing local and global climates has been simplification. In the EU, researchers and analysts first reduced the problem to greenhouse gas emissions, with carbon dioxide identified as the greatest concern. Then they proposed transportation and electricity generation as the easiest areas of interventions for mitigation. The production of agrofuels and biofuels became a major policy focus—a supranational problem requiring a global response. The intricacy of local variation was eliminated. Complex mathematics and increasingly sophisticated technology enabled the creation of globalised statistics and databases and the representation of the earth as an assembly of global grid cells.

The environmental history presented in this paper combines an analysis of the origins and development of globalised soil, land and terrain databases and models with an exercise in data literacy. It demonstrates confluences of socioeconomic and technological factors in the late twentieth and early twenty-first centuries. Digitisation stimulated globalised data manufacture and manipulation. Numbers stripped of their original context acquired identities of their own. From them, new databases were derived. Filtered or smoothed soil descriptions associated with large

86 For example, David Anderson and Richard Grove, eds, *Conservation in Africa: People, Policies and Practice* (Cambridge: Cambridge University Press, 1987); Richard H. Grove, *Green Imperialism: Colonial Expansion, Tropical Island Edens and the Origins of Environmentalism 1600–1860* (Cambridge: Cambridge University Press, 1995); W. M. Adams, *Wasting the Rain: Rivers, People and Planning in Africa* (Minneapolis, MN: University of Minnesota Press, 1992); Stephen Dovers, Ruth Edgecombe and Bill Guest, *South Africa's Environmental History: Cases and Comparisons* (Cape Town: David Philip Publishers, 2002; Athens: Ohio University Press, 2003); Thomas J. Bassett and Donald Crummey, eds, *African Savannas: Global Narratives and Local Knowledge of Environmental Change* (Oxford: James Currey; Portsmouth, NH: Heinemann, 2003); Kate B. Showers, *Imperial Gullies: Soil Erosion and Conservation in Lesotho* (Athens, OH: Ohio University Press, 2005); Edmund Burke III and Kenneth Pomeranz, eds, *The Environment and World History* (Berkeley, CA: University of California Press, 2009).

numbers of grid cells were presented as soil maps. These processes of simplification were required for harmonisation, and grid-cell construction served to erase many connections, as well as interactions, between soils and the ecosystems of which they are a fundamental and essential part.

Constructing globalised statistics, databases and models involves a continuous series of decisions about what to include and what to exclude. The creator's or creators' professional and personal perspectives guide these choices. The resulting, usually unacknowledged and largely invisible, biases become embedded in the structure of databases and models. Their effects can be amplified with further mathematical manipulation or inclusion in larger databases or models. Because of this, despite mathematical sophistication, neither databases nor models are objective or neutral. So-called science-based decision-making is, thus, culturally determined.

Exclusion extends far beyond choices about data set and model components. It is a form of social power. In its most aggressive form, exclusion is exercised in an effort to control, through denial of access, of existence or of significance. In a more benign form, exclusion simply prevents model users from having a broad perspective, from understanding complexity and from being able to design policy that can withstand the unexpected. Alternatives and local potential are eliminated.

Questions should be raised about what has been lost in the sequential processes of simplification fundamental to globalised database construction and analyses. Although facilitating further modelling, each mathematical manipulation increases the distance between the socially constructed world of global grid cells and the biogeophysical properties of specific places. Simplified values can produce technically robust analyses that are irrelevant at best—and dangerously misleading or completely false at worst. Policy formulated at the scale of simplified global grid cells is structurally incapable of engaging with the local scale at which responses can take place.

Data literacy, the critical analysis of statistics, databases and models, has practical and ethical significance. Environmental historians have documented past economic, social and environmental calamities created by ignorant interventions in societies and their environments. The potential for negative consequences to be not only repeated but magnified in scope exists when planning and project implementation are conceived of at a global scale. At the same time, urgently needed appropriate local responses will be obscured, if not denied. It is, therefore, crucial that possible consequences of policy and actions based on coarse-scale databases be examined in the context of local conditions before being mandated and implementation attempted in actual biogeophysical landscapes. If not, who will be responsible morally, as well as financially, for the negative environmental and socioeconomic consequences that policy formulated at the global grid cell will certainly create?

This text is taken from *International Review of Environmental History*,
Volume 5, Issue 1, 2019, edited by James Beattie, published 2019 by ANU Press,
The Australian National University, Canberra, Australia.

doi.org/10.22459/IREH.05.01.2019.05