Geoderma xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Past, present & future of information technology in pedometrics

David G. Rossiter*

ISRIC World Soil Information, PO Box 353, Wageningen 6700 AJ, The Netherlands
Section of Soil & Crop Sciences, Cornell University, 242 Emerson Hall, Ithaca, NY 14850, USA
School of Geography Science, Nanjing Normal University, 1 Wenyuan Road, Nanjing, Jiangsu 210046, China

ARTICLE INFO

Handling Editor: Dr. A.B. McBratney

Keywords:
Information technology
Pedometrics
Simulation models
Geostatistics
Numerical taxonomy
Soil-landscape evolution

ABSTRACT

Although pedometric approaches were taken as early as 1925, the post-WWII development of information technology radically transformed the possibilities for applying mathematical methods to soil science. The most significant development is the digital computer and associated developments in computer science. These allowed statistical inference on large pedometric datasets, e.g., numerical taxonomy of soils from the early 1960s and geostatistics from the mid-1970s, as well as simulation models of soil processes. By the time of the first Pedometrics conference in 1992 sufficient computing power was available for stochastic simulation and complex geostatistical procedures such as disjunctive kriging. In the intervening 25 years computing power has grown to almost magical proportions, allowing individual scientists to carry out complex procedures. A second development is the growth of networking. This facilitates collaboration among pedometricians, rapid communication with journals, collaborative programming and publication, and easy access to resources. A third development is the availability of on-line storage of large datasets, especially of open data, including GIS coverages and remotely-sensed images. This allows pedometricians working on geographic problems to integrate sources from multiple disciplines, most notably in digital soil mapping using a wide variety of covariates related to soil genesis. It also promotes the an open-source movement of collaborative development of computer programs useful for pedometrics. A fourth development is the wide range of digital sensors which provide data for pedometrics; sensors include spectroscopy, electromagnetic induction, and γ-ray detectors., connected to each other and to central data stores. A fifth development is wireless technology, including mobile computing and telephony, again greatly facilitating rapid and extensive data collection – in pedometrics, the more dense the data, the greater the analytical possibilities and the lower the uncertainty. A final development is a global navigation satellite system (e.g., GPS), making accurate georeference of field data a routine part of data collection, and thereby assuring the highest possible accuracy in maps made by predictive pedometric methods. As computer power increases, models can be correspondingly detailed. As sensor networks and remote sensing provide ever more abundant and spatiotemporally fine-resolution measurements, the challenge is to manage this information and maintain the link with pedologic and soil-landscape knowledge, within the context of societal needs for the results of pedometric analysis.

1. Introduction

Information technology (IT) is defined by the Oxford English Dictionary as "[t]he branch of technology concerned with the dissemination, processing, and storage of information, esp. by means of computers; abbreviated *IT*." This definition implies that algorithms and computational methods as such are not part of IT; rather, the technology to implement them is included in the concept. Acquisition of information (e.g., by digital sensor systems) is not included, only its processing, storage, dissemination once acquired; however since data must be acquired before it can be used in pedometrics, we include it in the present discussion. Information technology includes (1) data processing ("digital computers"); (2) digital data storage; (3) digital information sources; (4) computer operating systems; (5) computer programs; information standards; (6) communication & networks; and (7) digital sensors (satellite, air, field, lab.) and their associated digital infrastructure to convert raw sensor data into information products.

Pedometrics may be defined as "the application of mathematical and statistical methods in the study of the distribution, the characterization and the genesis of soils" (Heuvelink, 2003).

Thus the topic of this paper is the development of IT as it has affected pedometrics: how pedometricians work with IT, what aspects of

* Section of Soil & Crop Sciences, Cornell University, 242 Emerson Hall, Ithaca, NY 14850, USA.
E-mail addresses: david.rossiter@isric.org, d.g.rossiter@cornell.edu.

https://doi.org/10.1016/j.geoderma.2018.03.009
Received 3 November 2017; Received in revised form 2 March 2018; Accepted 12 March 2018
0016-7061/

Please cite this article as: Rossiter, D., Geoderma (2018), https://doi.org/10.1016/j.geoderma.2018.03.009

IT are most important to the work of pedometricians, how IT has changed the way pedometricians work and the topics we take up, and how it may affect our work in the near future. It does not consider the intellectual development of pedometrics, only how new possibilities in IT have facilitated advances in pedometrics.

Pedometricians are a sub-class of academic and research soil scientists, which are themselves a sub-class of academic and research earth or biological scientists, which are themselves a sub-class of research scientists in general. The use of IT thus can be explained in a hierarchy. All research scientists, including academic and research soil scientists, use IT for preparing and editing manuscripts and proposals (using text, mathematics and graphics), presenting scientific work to peers, funders and the public (perhaps also using animations), searching for and reading scientific results, designing research, and communicating and collaborating with peers. In addition, earth scientists use IT to navigate and collect data in the field using various sensor systems, to search for and use digital data on the Earth system, to analyze samples with computer-controlled or -assisted lab. instrumentation, and to organize data and results in databases and networked repositories. In addition, pedometricians use IT to write and execute pedometric algorithms (1) to summarize and manipulate sensor data; (2) to model soil properties and functions, spatio-temporal variability, and landscape evolution; (3) to predict and simulate using these models; (4) to design sampling schemes; and (5) to produce digital soil maps or other soil geographic databases.

2. Pedometrics timeline

The development of pedometrics up to the first Pedometrics conference in 1992 was reviewed by Webster (1994). From 1975 to 1983 an IUSS working group on soil information systems, reviewed by Burrough (1991), dealt with IT as it related to acquisition, storage and retrieval of map and point data. I reviewed the above-cited papers and all the papers in the special journal issues resulting from the Pedometrics conferences from 1992 (Wageningen) (De Gruijter et al., 1994), 1997 (Madison) (De Gruijter, 1999), 1999 (Sydney) (Odeh and McBratney, 2001), 2001 (Ghent) (Van Meirvenne and Goovaerts, 2003), 2003 (Reading) (Oliver and Lark, 2005), 2005 (Naples FL), 2007 (Tübingen), 2009 (Beijing) (Zhu et al., 2012), 2011 (Třešt, CZ) (Borůvka et al., 2014), 2013 (Nairobi) (Anonymous, 2016) and 2015 (Córdoba) (Vanwalleghem et al., 2018), as well as the Fuzzy Sets in Soil Science conference 1995 (de Gruijter et al., 1997).

Very few pedometrics-related papers in Geoderma, the European Journal of Soil Science (EJSS), the Soil Science Society of America Journal (SSSAJ), Mathematical {Geology, Geosciences} and similar journals describe the IT used (computer hardware, networks). Some mention the computer programs used, but all describe the algorithms, e.g., simulated annealing, sensitivity analysis, random forests, and generalized linear models, process models. From the nature of the algorithms and datasets we can infer something about the IT needed for the reported studies. If field or remote sensors were used for data acquisition, they are described.

3. History of IT use in pedometrics

A set of tables (see Supplementary material) showing selected milestones in various aspects of IT relevant for pedometrics is provided for context to the following discussion.

Going back to the earliest days of quantitative soil science, in 1910 Mercer and Hall (1911) studied the variability of wheat yield in an apparently uniform field, one of a series of uniformity trials popular at the time (Cochran, 1937). Although they did not measure the soil, they inferred that the substantial local variability in crop yield and grain/straw ratio could in part be attributed to local differences in soil nutrients. They computed numerical summaries, linear regression, and probable errors, all computed by hand. To simplify hand computations for linear regression they grouped observations, rather than use all 500 to fit the model. Fifteen years later Haines and Keen (1925a) carried out a series of dynometer experiments and produced a contour map of the drawbar pull required over a field, again by hand; this may be the first two-dimensional continuous soil property map. In a subsequent article (Haines and Keen, 1925b) they expanded this to a 3D solid model. In 1937 Youden and Mehlich (1937) computed what were effectively variograms of soil properties at spacings of a nested sampling design, along with a nested analysis of variance, again all by hand.

The first moves beyond hand computation were made possible by the development of the digital computer. An early pedometric application was numerical taxonomy applied to soils (Arkley, 1971; Bidwell and Hole, 1964; Hole and Hironaka, 1960; Rayner, 1966). In 1960 Hole and Hironaka (1960) were apparently computing taxonomic distances by hand, probably using a mechanical calculator. They constructed 3D models of similarity by cutting rods in lengths proportional to the dissimilarity indices and connecting them into a ball-and-stick figure. In 1964 Bidwell and Hole (1964) mention “electronic computers”, without specifying which, implying that matrices and dendrograms were computed on a mainframe computer. That paper includes a dendrogram, a graphical similarity matrix, and a 3D ordination plot; however these appear to have been drawn by hand based on the results of computations, not drawn by a computer graphics program. By 1971 Arkley (1971) mentions “large computer and multivariate statistics”, but does not specify what is “large” nor which computer was used.

Also during the 1960s, models of individual soil processes, such as heat and water transport, were developed; “consist[ing] mostly of analytical solutions of partial differential equations for well-defined soils and porous media, numerical solutions of single partial differential equations, or conceptual models that were solved with analog or digital computers” (Vereecken et al., 2016).

During the early development of soil information systems (1975–1983) computers came into wide use. Typical is the information system of STIBOKA (Dutch Soil Survey Institute) in 1977 (Bie and Schelling, 1978). This consisted of a CDC Cyber 72 mainframe computer, the NOS/BE operating system, the ANSI FORTRAN IV programming language, a relational database built by the National Environmental Research Council (UK), a Data General NOVA 1200 minicomputer (32K words) for graphics, a digitizer and plotter; 19” CRT display, a 14M word disk, a 9 track 800 b.p.i. magnetic tape, and teletype input. In 1984 this homebrew database was replaced with a commercial product (ORACLE) (Buurman and Sevink, 1995). This illustrates the increasing maturity of commercial IT offerings over this decade.

This was also the period where geostatistical approaches, primarily various forms of kriging, were first applied to pedometrics. A typical paper is by Yates et al. (1986) in which a FORTRAN IV program to perform disjunctive kriging is described. Their example dataset consisted of only 92 points. Although the authors did not discuss limitations of computing resources, the implication is that only small datasets could be handled by the IT available to them. Towards the end of this period the so-called “personal” computer, on the desktop of the individual pedometrician, became available (e.g., IBM PC 1982, Apple Macintosh 1984). Computer programs specifically for these were soon developed, for example the IDRISI GIS grid-based GIS, developed from 1987 (Eastman and Fulk, 1993). For more computationally-demanding applications, this era also saw the advent of desktop Unix workstations (e.g., Sun-1 1982), although these do not appear to have been used by pedometricians until about ten years later.

Towards the end of the 1980s, the development of more powerful computers and improved numerical methods (Press, 1986) integrated into subroutine packages (e.g., Dongarra, 1979) allowed the development of soil models that integrated different types of processes (physical, chemical, biological) and their interaction; these were applied at various scales from pore to global at different levels of process generalization (Vereecken et al., 2016).