Received: 4 March 2021 | Accepted: 17 June 2021

DOI: 10.1002/saj2.20296

Soil Science Society of America Journal

**2019 BRADY LECTURESHIP PAPER**

# Digital soil science and beyond

**Alexandre M. J.-C. Wadoux** | **Alex B. McBratney**

Sydney Institute of Agriculture & School of Life and Environmental Sciences, The Univ. of Sydney, Sydney, NSW, Australia

**Correspondence**
Alexandre M. J.-C. Wadoux, Sydney Institute of Agriculture & School of Life and Environmental Sciences, The Univ. of Sydney, Sydney, NSW, Australia.
Email: alexandre.wadoux@sydney.edu.au

Assigned to Associate Editor Kyungsoo Yoo.

Parts of this article were presented by Alex B. McBratney in a Nyle C. Brady Frontiers of Soil Science Lecture for the ASA–CSSA–SSSA International Annual Meeting in San Antonio, TX, on 11 Nov. 2019.

**Abstract**

Digital convergence is helping us to better understand and study the soil. Fixed and mobile sensors, and wireless communication systems aided by the internet produce cheap and abundant streams of digital soil data that can readily be used for modeling and information generation. Here, we explore the ways in which digital science and technology have affected soil science. We can call this digital soil science and define it as the study of the soil aided by the tools of the digital convergence. To some degree, all of our research and teaching had been enabled, enhanced, and expanded by the digital convergence. We outline how soil science has changed using illustrations of intellectual and technical developments enabled digitally. Digital soil sensors have been widely implemented, and new tools such as cell phones and applications, or metagenomics techniques are becoming available. There are also areas in soil science for which no major obstacles in the digital technologies exist, but which have not been thoroughly investigated—for example, to devise a truly digital soil field description or for building a formal digital quantitative system of soil classification. The soil science community will need to be alert to some of the dangers brought by digital convergence such as the lack of new theory and proprietary (black-box) soil prediction. Finally, we discuss a whole set of digital tools that will, or might, gain the stage in the immediate future and take a stab in the dark on what may lie over the horizon of digital soil science.

## 1 | INTRODUCTION

Stepping back in time 45 years, in 1976, a young scientist collected soil data from 101 soil profiles on a 2-km-long transect at Tillycorthie near Aberdeen, Scotland. The analysis of soil variation presented in McBratney and Webster (1981) using techniques of data transformation, principal component analysis, and computation of the sample variogram required a personal computer and about 0.1 MB of computer storage. While quite unremarkable for the present-day soil scientist, the use of digital computer technologies in 1981 was already a long way from other tools such as punched cards, a pre-electronic digital data storing and handling tool widely used by soil scientists (e.g., by Beckett et al., 1972). A few years before the study of McBratney and Webster (1981), the punched cards were themselves an improvement in comparison with analog data such as aerial photographs used for soil and land evaluation by Buringh (1954) or Webster and Beckett (1970), among others. Fast forward to the present day, scientists have embraced and nurtured the digital environment, made from binary ones and zeros instead of analog data that required human interpretation. One makes digital maps of the world using large (>10 Gb) electronic digital soil databases of hundred of thousands of soil profiles. These soil data are acquired rapidly by digital sensors and instruments. Data analysis is enabled

**Abbreviations:** IoT, Internet of Things; ISRIC, International Soil Reference and Information Centre; LIDAR, light detection and ranging; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; PCR, polymerase chain reaction; vis-NIR, visible–near-infrared; WRB, World Reference Base.

© 2021 The Authors. Soil Science Society of America Journal © 2021 Soil Science Society of America

by computers, digital imagery, and cloud processing. Today's digital data generation, storing, processing, and visualization has been brought to a level that was unimaginable just 45 years ago.

### 1.1 | What "digital" means

The Oxford English Dictionary defines digital as "relating to numerical digits and (later) their use in representing data in computing and electronics." Etymologically, digital comes from the Latin word *digitus*, which in Roman languages such as French refers to the finger. In Germanic languages such as English, the word digital is used in the sense of numerical: an encoded representation based on a finite set of discrete elements (Strasser & Edwards, 2017). As such, there is no divide before and after the advent of computer technologies because soil data need not be in electronic format to be digital; any record as numbers or text in articles and books were already digital before becoming electronic. The major innovation of these recent years came from analog-to-digital conversion. Imagery, maps, reading from spectrometers and sensors, and drawings had to be transformed to a digital representation, which, with the emerging resources in computing, enable data comparison, analysis, processing, and sharing. This analog-to-digital conversion is not finished. At the time of the writing, several institutions in the world, such as the Institut de Recherche pour le Développement in France, the ESDAC (European Soil Data Center) in Italy, or International Soil Reference and Information Centre (ISRIC)—World Soil Information in the Netherlands are still digitizing hand-drawn soil maps (Figure 1) and their collection of analog soil data.

### 1.2 | Getting more information more cheaply

Besides the analog-to-digital conversion, sensors and instruments generating digital data have gradually replaced analog instruments which require human reading. The standard tensiometer measuring soil moisture content with a vacuum gauge has an analog display, whereas more recent ones have digital outputs and can be read by cell phone apps and computers. Instruments producing analog outputs are equipped with an accessory to instantly digitize the record. For example, recent soil spectrometers have embedded analog-to-digital converters to translate the measuring signals emitted by the detector into a number of discrete elements readable by a computer. Thus, one of the main features of this analog-to-digital conversion is the opportunity to produce more data more cheaply, which can readily be used by computers and information technologies. The digital convergence has continued with the internet, and more recently with the development of new tools to handle and process large amount of electronic data. Digital soil science is, by a semantic shift, the study of the soil aided by the tools of the digital convergence. How the technologies have facilitated advances in soil science and pedometrics is discussed in Rossiter (2018) and how the scientific methods have evolved with the recent technological change is reviewed by Wadoux, Román-Dobarco,and McBratney (2021).

**Core Ideas**

- Many aspects of soil science have been strengthened by the digital convergence.
- We outline how soil science has changed with the digital.
- The dangers and immediate future of digital soil science are discussed.
- Scenarios on the far future of digital soil science are proposed.

### 1.3 | Chronology

The major events of the digital convergence in soil science—namely, convergence of digital data, electronic databases stored in computers, the internet, and new tools to handle

FIGURE 1 Digitization of a legacy soil map at International Soil Reference and Information Centre (ISRIC)—World Soil Information in the Netherlands in 2021. The operator places the soil map in a high-resolution scanner. A digital copy of the map is then rendered in the computer and manually georeferenced. The map is stored in a raster graphic format, which enable computer processing and manipulation in a geographic information system (e.g., to recreate mapping units). Picture courtesy of ISRIC—World Soil Information

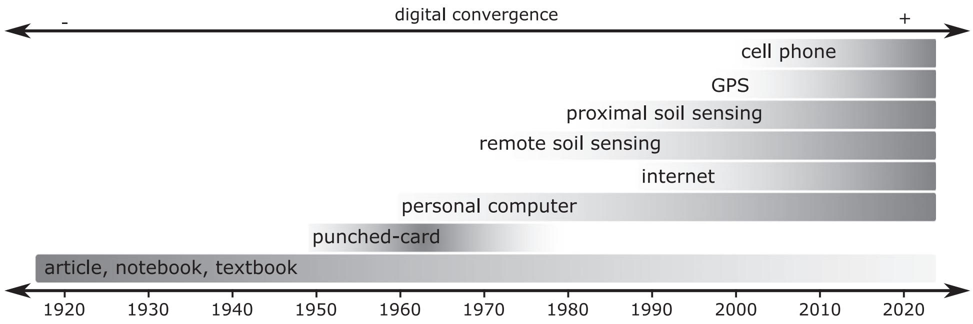


**FIGURE 2** Simplified chronology of the major events of the digital convergence in soil science

and process large amount of electronic data—have followed the developments of the digital revolution in other sciences and in society. We outline a brief chronology of the main events (Figure 2). In keeping with the previous paragraph, digital data have existed long before the computer age in pre-electronic digital form stored in cabinets or research institutions as numbers and text in articles, notebooks, or textbooks. The punched-card system was adopted in soil science from the 1950s (e.g., by Muir & Hardie, 1962; Wischmeier, 1955), but it was the general purpose computer and the conversion of pre-electronic digital soil databases into digital ones that brought a new tool for the soil scientist in the 1960s. With the computer, the dramatic expansion of the digital soil science began, aided by the internet in the 1990s and the widespread use of digital remote, airborne (1980s), or proximal (1990s) sensors and instruments. One of the first instruments proposing a digital signal were geophysical sensors EM31 in the 1970s. The global positioning systems (GPS) became widely available to civilians by the 2000s. Finally, cell phones and the high-speed internet became available to a large audience from the 2000s onward.

### 1.4 | Objectives of the paper

In this paper, we shall explore how digital science and technology has affected our science—the study of the soil. To some degree, all of our research and teaching had been enabled, enhanced, and expanded by the digital convergence. Indeed, the understanding of all aspects of "soil" has been strengthened and intensified by digital data, technologies, and approaches. We will outline how soil science has changed using illustrations of intellectual and technical developments enabled by the digital convergence. The examples originate largely from work in pedometrics, digital soil mapping, digital agriculture, and soil security. We further discuss actual failures and potential dangers of digital soil science. Finally, we will speculate on what might evolve over the next decade and take a stab in the dark to speculate on what may lie over the horizon.

## 2 | WHAT HAS DIGITAL FACILITATED? MAJOR SUCCESSES

### 2.1 | Remote soil sensing

Soil scientists have used remote sensing since the 1920s to segment the soil-landscape into homogeneous areas with aerial photographs (e.g., Bushnell, 1929), but it was the full-scale digitization that made remote sensing images available for computer processing or statistical inference. Manual digitization of early analog Landsat imagery began in the 1970s (Rogers et al., 1975). Later, analog sensor data were immediately digitized on-board the spaceborne or airborne with embedded analog-to-digital devices. Early studies of remote digital sensing of soils used airborne multispectral data or passive (radiometry) and active (radar) microwave techniques (see, for example, Kristof et al., 1973; Schmugge, 1978).

For nearly four decades, soil scientists have taken advantage of the constant stream of digital data brought by remote sensing. A large number of passive and active remote sensors were developed (Wulf et al., 2015): optical multispectral, spectroscopy, thermal, or microwaves, but also radar and light detection and ranging (LIDAR) sensors such as synthetic aperture radar (SAR) systems. The spectral range available for soil scientists have broadened, and so the applications. Remote sensing can characterize a large range of soil properties: minerals, texture, moisture, carbon, salinity, carbonates, and contaminants. Palacios-Orueta and Ustin (1998) determined topsoil iron and organic matter contents from visible and infrared remote sensing data in two watersheds of the Santa Monica Mountains, USA. Dehaan and Taylor (2003) estimated soil salinity with hyperspectral imagery to map irrigation-induced salinization in the Murray Basin, Australia. Mulder et al. (2011) went on to recognize three main components that remote soil sensing data support: (a) the segmentation of the landscape into homogeneous and mutually contrasting soil-landscape units, (b) the estimation of soil properties using physical or empirical models, and (c) the interpolation of soil point information as a secondary data source.

The analysis of large streams of digital remote sensing data has been facilitated by digital tools. Computer processing soon replaced hand computation of analog images. For example, Palacios-Orueta and Ustin (1998) computed the depth of the band and performed principal component analysis on the spectra, whereas Vaudour et al. (2019) used a large stack of Sentinel 2 data, partial least squares regression and variogram analysis for mapping several soil properties. Among recent digital developments, collaborative platforms for preprocessed data sharing emerged. ScienceEarth (Xu et al., 2020) or Google Earth Engine (Yu & Gong, 2012) are examples of cloud-based computing environments. Other interfaces administer large spatio-temporal remote sensing datasets, for which daily production exceeds 10 Tb—for example the DataCube (Lewis et al., 2017), which has recently been adapted to the open-source R programming environment (Appel & Pebesma, 2019).

## 2.2 | Proximal soil sensing

In the 1900s, a large number of studies and patents applied the sensing principles to the study and mapping of the subsurface. Schlumberger (1920) details several electrical prospecting methods for mineral detection and mapping. As early as 1938, a review by Rust (1938) declared that about 100 articles were being published each year on electrical prospecting. Using these principles, a study by Haines and Keen (1925) is presumably (McBratney & Minasny, 2010) one of the first proximal soil sensing studies. The authors used an on-the-go soil sensor and made a high-resolution analog map of the soil mechanical resistance of a field at the Rothamsted Experimental Station in England. As the timeframe suggests, the maps were hand drawn and the data from the sensors required human reading. Here, too, the last decades have witnessed an explosion in the variety of proximal and digital soil sensors and techniques: electrical resistivity, laser diffraction, visible and infrared sensors, digital cameras, hyperspectral scanner, X-rays, micro- or radio waves, gravity, and more (Viscarra Rossel et al., 2011). Proximal soil sensing can be loosely defined as the measure of soil characteristics using sensors at a distance from the soil no greater than 2 m. It has provided abundant and cheap digital data for soil scientists and has fueled the development of precision agriculture and the subdiscipline of digital soil morphometrics.

The most significant developments in proximal soil sensing were led by progress in digital soil spectroscopy, primarily in the visible and infrared range of the spectrum. Research on laboratory and field spectroscopy began in the 1970s (see, for example, White, 1971), but it is only in the past 30 yr, coinciding with the development of digital sensors and tools, that soil spectroscopy has become an active area of research. Soil scientists in the 1980s used laboratory spectrometers to rapidly measure the chemical composition of the soil, focusing mainly on texture, mineralogy, and soil organic matter. The use of a portable spectrometer is described from the 1990s (Shonk et al., 1991), followed a decade later by on-the-go soil properties estimates with field spectrometers (Stenberg et al., 2010). In addition to the abundant amount of digital data produced by spectrometers, digital tools to store and handle these data have also developed. Soil spectral libraries are digital data frames of soil infrared spectra combined largely with wet chemistry measurements. Digital convergence continued to trigger developments with the internet and permitted a global collaboration in 2016, achieved with the publication of a electronic global soil spectral library (Viscarra Rossel et al., 2016) composed of 23,631 digital spectra. As a corollary to the increase in the availability and diffusion of digital spectra, a new field of research called chemometrics developed in the use of data analytic for spectroscopic data, with adaptation of these tools in open source programming languages (e.g., Wadoux, Malone, et al., 2021).

Maturity in the use of various digital sensors in soil science has led to the integration of multiple proximal sensors into a single system. Using complementary sets of sensors has several advantages in estimating various soil properties in the field; in particular it increases the range of the electromagnetic spectrum covered and it enables estimating soil properties with more confidence than when using a single sensors. An example is the study of Taylor et al. (2006), which reports on the development of a multisensor platform with two electromagnetic induction instruments, passive gamma, electrical resistivity, and pH sensor. On-the-go soil measurements with platforms is enabled by real-time kinematic GPS. Digital GPS data and associated receivers are imperative for proximal soil sensing and to combine the platform measurements with a set of low-cost, digital, and accurate environmental information such as elevation and slope at the measurement point (Viscarra Rossel et al., 2011). An example of a multisensor platform is shown in Figure 3.

## 2.3 | Digital soil mapping

Early soil maps presented in analog form often result from a compilation of analog data and data from field surveys that are digital in form (e.g., numbers serving to define the content of the soil map units). The development of the computer in the 1960s and numerical processing made possible digital representation of soil maps, first by stepped boundaries and later with curved lines (Legros, 2006). Digital data stored in punched-cards, punched tapes, and magnetic tape were translated to an electronic digital format to being available for numerical analysis. In the 1970s, for example, Webster and Burrough (1972) collected 84 soil samples on a grid for an area of north Berkshire in the United Kingdom.


**FIGURE 3** A multisensor soil sensing platform composed of a RSX-1 gamma spectrometer (Radiation Solutions) and an electromagnetic induction instrument (Dualem21; Dualem). The sensors are mounted on a field vehicle equipped with a high-precision differential GPS (DGPS) to sense the soil of a vineyard of the Pokolbin region in the lower Hunter Valley of New South Wales, Australia

Principal component analysis on 17 soil properties and mapping of the first component revealed agreement of the soil spatial variation with an existing soil series map. From this early example, progress has been considerable with the soil information systems, relational database management systems, and remote sensing in the 1970s, and the geographic information systems and geostatistics in the 1980s. In the 2000s, the explosion of information technology and digital soil data available for digital soil mapping, denoised GPS systems coupled with new tools in statistics and data mining, and cloud processing have made digital soil mapping a very successful subdiscipline of soil science (Minasny & McBratney, 2016).

The current trend in digital soil mapping activities illustrates the extreme reliance on the digital. For example, Stockmann et al. (2015) made a global-scale and spatio-temporal assessment of topsoil organic carbon at 500-m resolution. The authors used time series of remote sensing images stored in the Google Earth Engine platform and preformed cloud processing on 15 million (land) pixels per year (i.e., around 238,240 million pixels between 2001 and 2016). This is in line with Lagacherie and McBratney (2006), who evoked Moore's law (Moore, 1965). They predicted two decades ago that increase in computer power would lead to an exponential growth in the number of pixel that future digital soil mapping projects could tackle. Minasny and McBratney (2016) revisited this model and found that the number of pixel could double every year. In fact, we may argue that we have far exceeded these predictions. Chaney et al. (2019) generated a probabilistic soil property maps of the United States at about 30-m resolution. With a high performance distributed cloud computer, it took only 5 h to predict 9 billion grid cells.

With the digital tools, applications of digital soil mapping are constantly expanding. For example, Lagacherie et al. (1995) used data from a soil survey in a reference area for mapping in an independent, wider area. Data mining was used by Bui and Moran (2003) to fill gaps in soil survey over large areas in the Murray–Darling basin in Australia. In this study, environmental covariates are combined with existing soil maps to model and predict soil types. The use of data mining and machine learning for digital soil mapping has attracted much attention in the past two decades; it has been covered in a recent review by Wadoux et al. (2020). Other examples of digital soil mapping studies are Häring et al. (2012), who used rules from a calibrated model to disaggregate soil map units into soil series, whereas Kempen et al. (2009) used multinomial logistic regression to model the relationship between environmental covariates and soil groups of a legacy soil map and updated the map by estimating the probability of occurrence of major soil group on a fine grid. Review of digital soil mapping studies are available in Lagacherie and McBratney (2006) and Minasny and McBratney (2016) .

## 2.4 | Soil microbial characterization

In the mid-19th century, Pasteur developed methods of isolation and cultivation of microorganisms. On this foundation, early studies on soil microbial diversity used culture-based methods to isolate and purify soil bacterial species and classify the isolates based on phenotypical characteristics (Tate, 2020). The species composition between sites and the diversity of a community is quantified with numerical methods such as cluster analysis or through computation of diversity indices. No doubt that more complete analyses are now permitted with the digital and the computer-based numerical analysis. Procedures to estimate soil biological diversity using surrogates (e.g., community-level physiological profiling and phospholipid fatty acid analyses) likewise benefited from the digital convergence. As for most present-day sensors, gas chromatography used in phospholipid fatty acid analysis now has embedded analog-to-digital signal converters, and digital data of community abundance are treated with multivariate techniques such as principal component analysis. For example, Kelly et al. (1999) used principal component analysis to evaluate the effect of sewage sludge on soil heavy metal concentrations and soil microbial communities. However, it is with the advent of molecular methods based mostly on DNA analyses coupled with developments in bioinformatics that analysis of digital data produced by sequencing techniques were made possible.

Target genes (e.g., 16S ribosomal RNA [rRNA]) of DNA extracted directly from soil samples (i.e., community DNA) are amplified by polymerase chain reaction (PCR) based techniques. The recent development of digital high-throughput amplicon sequencing techniques has enabled new approaches for the profiling of microbial communities. Metabarcoding allows to determine which microbial species are present in a soil sample by (a) targeting specific sequence of the DNA (i.e., barcode), (b) sequencing the corresponding DNA amplicons, and (c) bioinformatics analysis of the sequences. High-throughput sequencing is permitted by new digital platforms called next-generation sequencing (NGS) technologies, such as Illumina (Caporaso et al., 2012) or Ion Torrent (Whiteley et al., 2012). The Illumina HiSeq2000, for example, generates more than 50 Gb of digital data per day, which is equivalent in a 10-d period to 1.6 billion 100-bp-end reads (Caporaso et al., 2012). The incredible amount of digital data generated by these NGS technologies has triggered the development of bioinformatics tools and software to analyze these large datasets. Software packages such as QIIME (Caporaso et al., 2010) or mothur (Schloss et al., 2009) are specifically designed for the analysis of 16S rRNA gene amplicon libraries from NGS technologies. With the internet and availability of electronic data storage, sequences can also be compared to curated taxonomic reference libraries (e.g., by BLAST [basic local alignment search tool] search in the National Center for Biotechnology Information [NCBI] Nucleotide Database) to identify the soil biodiversity at a site.

The development of metagenomics techniques has also been facilitated by the digital. Instead of targeting specific fragments of the genome, metagenomics aims at sequencing the genomes of a population of microorganisms, without PCR amplification. However, metagenomics produces massive amounts of digital data, amounts that increase as the sequencing technologies change and the digital tools to handle them become more efficient. A typical microbiome study generates several gigabytes of digital data, but the amount of data produced can increase dramatically as the sequencing depth increases. Analysis of these data is done on computer clusters and requires extensive bioinformatics expertise. Sequencing a soil metagenome is ongoing research, but initiatives such as the TerraGenome international sequencing consortium (Vogel et al., 2009) attempt to sequence and annotate a reference soil metagenome. The tools to reach such objectives are international collaborations and open-system electronic data management and sharing through internet web pages and platforms.

## 2.5 | Cell phone and soil applications

The immediate popularity of cell (mobile) phones from the 2000s onward and the integration of digital cameras into them have motivated soil scientists to devise new ways of characterizing the soil. Since 2010, most cell phones have embedded high-resolution digital cameras and GPS positioning. Using principles of remote and proximal soil sensing, several authors have used digital images of soil recorded under controlled, indoor conditions (e.g., Levin et al., 2005) to estimate soil properties such as soil organic carbon. Estimation of the soil properties relies on the R-G-B color information of the image pixels. Applying these principles, digital images from cell phones have been used to estimate soil color by Gómez-Robledo et al. (2013) and by Stiglitz et al. (2016) with an additional sensor. Digital images of 8 megapixels are stored on the phone with the JPEG compression format, and computing is made by the phone processor. Han et al. (2016) used the soil color recorded by the camera for near real-time classification of the soil profile with machine vision. This was taken one step further by Aitkenhead et al. (2016) by using the connection between the app and a server-side database to estimate multiple soil properties with cloud processing.

The principle of using cell phones for a range of rapid soil analyses was adapted into a relatively small number of applications making use of high-resolution digital cameras, information transfer, and cloud processing. Applications have mostly been used for fast and easy data collection—for example, Flynn et al. (2020) evaluate the SLAKES app. This application quantifies aggregate stability with a simple experiment by recording images of soil aggregates as they disaggregate. More recently, Golicz et al. (2020) developed a phone application for soil nutrient analysis by the using the phone as a portable reflectometer. They showed that the applications could accurately estimate nitrate concentration, but that phosphorous estimation was more complicated.

Cell phone software applications serve as an interface between science and citizen science, by enabling crowdsourcing of soil data by nonspecialists. The mySoil application (Shelley, 2012) developed by the British Geological Survey is intended for the nonexpert user of soil information. Based on the postcode or the GPS location, the user access simple soil information presented in a user-friendly way. The public is also invited to submit soil surface observations such as texture and description. Another, perhaps more successful, UK initiative is the Open Air Laboratories (OPAL; Lakeman-Fraser et al., 2016) project, in which the public is invited to participate actively to earthworm data collection with a clear field guide and documentation. They also allow the nonspecialists to obtain custom information based on the

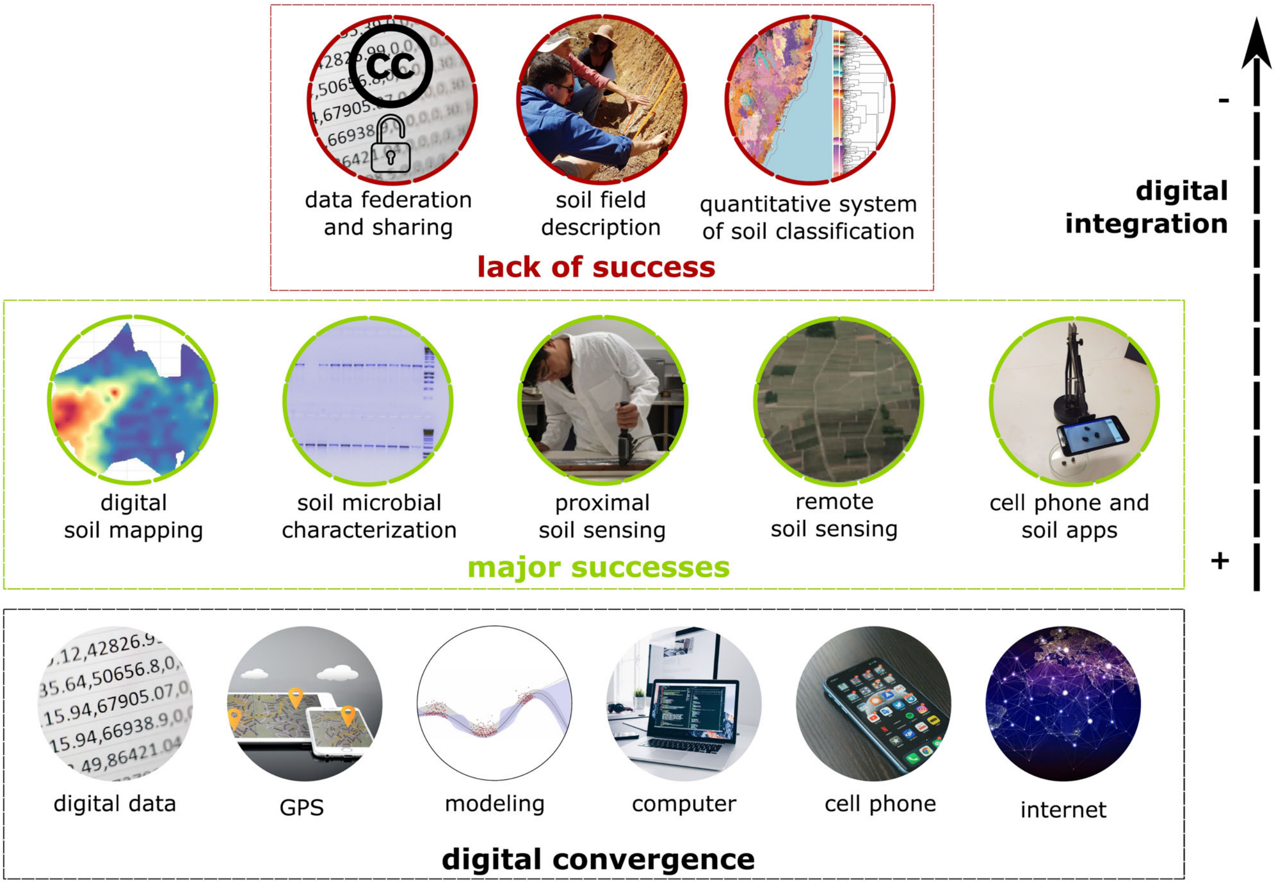


**FIGURE 4** Various aspects of soil science have been enabled, enhanced, and expanded by the digital convergence: namely, digital soil mapping, soil microbial characterization, proximal and remote soil sensing, and soil applications. Lack of an efficient digital integration in data federation and sharing, soil field description, and quantitative system of soil classification has, conversely, hindered intellectual and technical developments despite that no obvious obstacles from existing digital technologies exist

phone GPS location. For example, the purpose of the Soil-Info App of ISRIC (Hengl & Mendes de Jesus, 2015) is to provide access to soil data of the SoilGrids project. It also supports on-field digitization of soil profile data. These initiatives, enabled by the digital, facilitate public engagement and education around soil science and are reinforcing the feeling of empowerment and commitment of public in soil protection.

## 3 | LACK OF SUCCESS

Digital convergence has brought undeniable new capacity to analyze and study the soil. There remain some areas, however, where progress aided by the digital can be made, and for which we have the capabilities and no obvious obstacles from existing digital technologies (Figure 4). These present-day failures of digital soil science are discussed below.

### 3.1 | Data federation and sharing

In the past, soil data storage and compilation have been initiated with much enthusiasm since the 1980s (e.g., Ragg, 1977; Rudeforth, 1975), with a number of successful attempts in data collection and harmonization for the creation of local, regional, and global (Jian et al., 2020) digital databases. Bouma et al. (1999) discussed the need for soil databases in the context of precision agriculture: building a database represents a one-time, long-term investment valid for many years and is supported by recent information technologies.

Digital convergence indeed enables large-scale initiatives driven by soil database collection and harmonization. Soil scientists are also increasingly generating streams of data—for example, with proximal soil sensing techniques. Individual soil scientists, however, are not systematically placing their data into digital soil databases, either because of perceived impediments to data sharing (e.g., fear of misuse of data, perceived legal constraints) or because of unclear benefits that data sharing may bring. Lack of time and funding is one of the main reasons scientists do not share their data electronically, but they also recognize that their ability to answer scientific questions is hampered by the difficulty to access data produced by others (Tenopir et al., 2011). In Australia, the integration of new and legacy soil data is a current priority of the National Soil Research Development and Extension Strategy (Robinson et al., 2019). These efforts build on existing soil information systems and recent national open access policies to develop a digital spatial data infrastructure via a web service. Access to data produced by the public and private sectors is still conditioned to appropriate data licensing agreement. Recent developments in distributed ledger technologies, particularly the blockchain, are opening the way for a significant new (bottom-up) approach in data sharing, in which the data are not centralized into a single institution, but distributed through an online distributed database. Such a digital system has been simulated by Padarian and McBratney (2020) to build a global soil spectral library, with characteristics such as decentralized data sharing and governance. Clearly, the appropriate digital tools are available for soil scientists to put their data into soil national or international digital soil databases.

The next challenge concerns the further exploitation of data obtained by previous studies. Soil scientists need to capitalize on previous data collection, either by consulting databases to gauge the result of local studies, or to determine the possibility of generalizing some findings. There exists, for example, no unique database for metabarcoding data (Orgiazzi et al., 2015). Existing databases (e.g., NCBI) focus on DNA sequences with basic metadata, but the reuse of these data is limited to visualizing sequence variation among taxa. Orgiazzi et al. (2015) advocate the creation of DNA sequence data linked to soil properties and environmental information and metadata. This would enable mapping and further use of existing data for a purpose other than that which they were collected. They also warn against the loss of data of potential interest (e.g., the sequences identified as nonmicrobial). As our knowledge increases and technologies advance, storing these unused data into digital repositories may prove valuable in the future. This was also promoted in proximal soil sensing by Wadoux, Malone, et al. (2021): soil spectral data might reveal in the future soil information currently disregarded, such as microbial biomass carbon, polluting chemicals, and microplastics.

## 3.2 | Adopting digital soil field description

We have been slow to adopt digital technologies for data generation. This is particularly true for everyday field soil description which essentially uses the same technology as in the 1950s. For example, the FAO Guideline for soil description (Jahn et al., 2006) emphasizes descriptive information and qualitative estimates of soil properties, by no means using the tools from the digital convergence, with the exception of the GPS to obtain the geographic location and elevation data. In Australia, the second edition of the guidelines for surveying soil and land resources aimed at updating previous soil survey that were based on a logic that predated the computer (McKenzie et al., 2008), with, among others, chapters on sensing, pedometrics, and uncertainty. However, the authors noted that the adoption of digital technologies for soil measurement has not proceeded at the same pace than in other fields (e.g., precision agriculture). Soil scientists are lagging behind in terms of in-field digital technologies applied to field soil description. There is no widely adopted digital equivalent to the Munsell soil color chart for soil color analysis or to the 10% HCl effervescence test for evaluating calcium carbonate. Current proximal soil sensing techniques have changed the practices in soil measurements dramatically, and we need to apply these powerful digital technologies, coupled with new internet-based and connected data platforms and cell devices.

No single sensor can quantify all properties accurately. Currently large portions of the electromagnetic spectrum can be recorded in situ with portable devices, but their combination with other sensors and techniques is still underexploited. The field of digital soil morphometrics (Hartemink & Minasny, 2014), which study the soil profile as a basic entity with a range of digital tools and techniques, has brought many digital technologies to the field. We have enough technology to estimate quantitatively nearly all properties from the usual soil field description guideline of Jahn et al. (2006). Jones and McBratney (2016) argued in this sense and recommended to search for new technologies, in particular the noninvasive ones. Noninvasive sensors such as ground-penetrating radar and electromagnetic induction are already proven efficient in describing a set of soil properties without digging a soil pit. As for digital soil morphometrics, the soil field description needs to leverage the opportunity of the digital. Data fusion from multiple sensors coupled with connected soil inference systems (McBratney et al., 2002) will provide the most useful information. The great power in the near future will come from the wide adoption of easy-to-follow soil field description procedures, perhaps residing in cell phone apps, but the potential for fusing different sensors and digital techniques for soil field description are myriad.

## 3.3 | A formal digital quantitative system of soil classification

Soil classifications have been derived for a century or more. They combine our current understanding of soil processes and formation coupled with field observation and soil measurement. In the 1960s, the advent of the computer led soil scientists to study numerical classification, which was developing quickly in biological systematics. It also brought new tools that were almost immediately applied to create (quantitative) numerical soil classification from measurement of soil properties, in place of the existing qualitative classifications. Despite a major leap forward with the advent of computer and later various digital technologies, research on numerical soil classifications have not advanced much in the last four decades.

Existing computer statistical techniques of data analysis need to be applied, digital information systems to be used and sensing technologies to be adopted. Soil scientists are familiar with many of the computational and statistical methods of data analysis such as calculation of similarity and taxonomic distances. We should investigate how these numerical methods could serve as basis to a unified quantitative system for soil classification that enable transfer between existing numerical classifications and allocation to new individuals to the classification. Current classifications have all been developed to fulfill specific applications, but no unified system has been developed. There is also enough technology to think of a unified global soil classification providing a unbiased basis for merging the existing local and regional soil classification systems. The two existing global soil classification systems (World Reference Base [WRB] and Soil Taxonomy), have no direct linkage, and there are several problems in their application to obtain effective local-scale soil information (Hughes et al., 2017).

Ideally, we would be able to collect a large amount of data from a digital soil information system, either profile data of spatially continuous soil property maps, and apply to these data some numerical classification algorithms. Unsupervised classification with *k*-means is an example of such technique, but other techniques that include fuzziness in the classification could also be used. The numerical techniques and technologies have since long been used by soil scientists. We must also investigate how soil sensing technologies might be useful. Current sensing techniques have been successful to retrieve various soil properties in near real time (see also Section 2.2), and despite that they appear underexploited for soil field description (see Section 3.2), they might be useful for near-real-time allocation of soil individuals to a numerical soil classification. Sensing data quantify all attributes of interest. The great power in the coming years will come from the use of portable soil sensing devices coupled with a digital approach linking the recorded data to online, cloud-based data analysis tools. For example, on-the-go soil sensing data may be sent to a server and soil property estimates sent back to the user after prediction by precalibrated statistical or data mining models.

The time component has also been disregarded in current numerical soil classifications. In the WRB soil classification, for example, time is included indirectly through inclusion of the pedogenetic processes. The reason for not including time in most classification is the difficulty to obtain accurate information on past soil forming factors that led to the current soil spatial variability. The use of online platforms such as Google Earth Engine or DataCube and long-term digital soil evolution models may provide useful information for soil properties evolving fast (e.g., soil organic matter). Long-term records of remote sensing images are now available, which can be used as proxy for soil formation and evolution. Similarly, soil evolution and earth system models may provide digital information to efficiently incorporate the time dimension into classification. Including time in numerical soil classification is, in reality, still a long way off, and currently no model or temporal dataset is precise enough or has sufficient coverage to reconstruct soil past forming factors.

## 4 | THE DANGERS

### 4.1 | Lack of new theory

The foremost danger with the rise in digital information and increase in computer resources is to ensure that soil science studies do not become supply driven (Rossiter, 2018), to the detriment of theoretical studies. Digital data and tools are a great aid in facilitating research in soil science (we gave example of such successes in Section 2), but the danger might come from the temptation of using these technologies as an end in themselves. The digital can help answering the questions framed within theories; perhaps also the vast amount of data allows patterns to emerge that can be used to generate hypotheses. We have highlighted elsewhere (Wadoux & McBratney, 2021) how in digital soil mapping studies, the hypothesis can be the useful end point of the research in which the digital data are not used to corroborate a hypothesis, but to suggest possible explanations to a phenomenon. Wadoux, Román-Dobarco, and McBratney (2021) also discussed data-driven soil science, which relies on (digital) data to generate soil science knowledge. As the authors put it: theory-free analysis of data does not hold long in soil science. To be useful, the digital needs to go hand in hand with theoretical developments, because the digital helps to corroborate and refine the vision of the soil expressed by the theory, within a scientific approach. Accumulating studies on the latest digital developments may not serve this purpose and should not drive the soil science agenda. There is no end to digital development. There is reason to fear that the gap between digital soil science studies and theoretical studies will grow bigger. This will question the validity of future digital soil science studies, which may well mismatch real-world processes, if not grounded in any theory.

## 4.2 | Lack of soil science knowledge

In analyzing the current state of soil science, both in undergraduate and graduates programs and in scientific publications, it is clear that the digital is triggering a loss of traditional soil science expertise and knowledge. Philip (1991) lamented the lack of laboratory and field skills of young researchers in the 1990s, making them “blissfully unaware” of their inadequacy for soil science research. He attributed this loss of expertise to the development of computer modeling in research activities of students instead of more expensive studies involving investigation of the physical world. Hudson (1992) puts it: it takes 2–3 yr for a new field scientist to internalize the soil-landscape paradigm. More recently, Lobry de Bruyn et al. (2017) termed graduates in soil science as “distant and removed, metaphorically and sometimes literally, from landscapes.” This has serious consequences. First, it affects critical soil-related problem-solving skills. Second, and perhaps more worryingly, the digital causes an increasing separation between simulation studies and the real world they are supposed to represent. The loss of soil science expertise has several causes such as the emphasis on applied rather than basic research and the general decline in funding for soil science. Each of these causes has facilitated the development of the digital. The digital is said to be cheap, whereas field and laboratory work is expensive and time consuming. Relying only on the digital to the detriment of other forms of research is increasing separation between the digital soil science studies and the real world it is supposed to represent. Anyone with a computer, easy-to-access digital soil databases, and user-friendly software resources may produce decent results (Bouma, 2010), but expertise is required to flag these results when they are misleading. Another risk is to blindly rely on digital computer models not evaluated in the physical world, models that can be wrongfully interpreted as reality by users. The lack of soil science expertise facilitated by the digital is a potential source of unsound practices.

## 4.3 | Proprietary soil prediction

We refer to proprietary soil predictions as the acquisition of soil information from closed-source (free or commercialized) software (Hengl et al., 2018). With the increasing share of the digital in soil science, the use of software and complex numerical analysis is becoming routine. For example, spectrometers have usually embedded software for preprocessing and multivariate statistical analysis of the digital spectra. Software implementations are useful because they provide the majority with tools to render complex numerical analysis of digital soil data practicable, but the use of close-source software is problematic. Close-source software hide to the users the code and workflow that generated the predictions. As a result, in the treatment of digital soil data, and in particular in company-owned close-source software, very little transparency on the workflow is possible, and potential discrepancies between software cannot be readily explained. Such software are commonly referred to as “black boxes” and might be covered by patents on the workflow or the way they handle the digital soil data. Reliability of the software is also called into question, as uncertainty is seldom provided. For example, the so-called “Lab-in-the-Box” (AgroCares) is an on-the-spot soil testing instrument capable of predicting a large set of soil physical and chemical properties with a single infrared scan of a soil sample. What preprocessing is applied to the spectra and what workflow and regression models are applied to the spectra are undisclosed. The instrument is connected to an user-friendly cell phone application that provides users with the results of the soil scan. Ironically, this commercialized black-box system is made possible by the effort of many providers of noncopyrighted digital soil data, such as research institutions, individual researchers, and nonprofit organizations, and the creators of digital soil database, on which complex machine learning models can be precalibrated and sold.

## 4.4 | Lack of new data generation

It has been argued that with the digital, traditional soil inventory programs are coming into an end and primary data collection is meant to decline. In most Western countries, the liberalization of the economies since the 1970s and the short-term research programs are compromising the collection of new soil data. In Australia, for example, the number of soil profile description collected per year has dropped since the 1990s (Biggs et al., 2018). A similar trend is observed in other countries. With the digital, and in particular the advent or remote sensing methods and statistical modeling since the 1970s, providing spatially explicit estimates of soil properties, many areas of soil science are now “done from the desk,” and verification is only occasionally performed in the field (Philip, 1991). Projects involving the digital and digital data collect few primary soil data, and thus appear cheaper and to provide faster results than projects involving new data collection. Paraphrasing Thomas (1992), dollars are directed towards number generators rather than data takers. Digital projects may appear counterproductive because they are often done at the expense of new data collection (Basher, 1997), which in the long term will result in a decrease of the quality of the digital information and in poor return of investment for the funding agencies and national soil survey programs. This was also highlighted elsewhere (Ibanez, 2006): the digital cannot operate indefinitely only on the basis of inferred data (e.g., proximal or remote soil sensing). We need more field data to improve the efficiency of the digital technologies and electronic computerized simulations. In this digital soil science, the danger is to contribute only in collecting existing digital soil data, sharing them and generating new numbers without participating to the updating of soil inventories.

### 4.5 | Doing too much with too little

As a consequence to the lack of soil data collection, digital soil science raises concerns on whether we are tempting to do too much with too little. The current style is to use digital techniques, computer simulation aided by digital soil databases to generate empirical estimate—to generate numbers using, for example, pedotransfer functions or soil mapping models. Indeed, it is fashionable, and the capabilities to generate numbers, simulated estimates of soil properties or attributes, appear endless. Minasny and McBratney (2016) report on the increase in digital soil mapping resolution; we are producing soil maps that are always increasing in spatial resolution. In Australia, maps of soil properties were produced at 250-m resolution in 2005, 100-m resolution in 2011, and 30-m resolution in 2020, but the number of new primary soil data collected to support these increases did not follow the same trend via increased sampling density. Several authors have reported the lack of new soil data collection in Australia for the same period (see also Biggs et al., 2018, or Section 4.4). Another example are the soil property maps of Africa, produced at a spatial resolution of 250 m in 2015, and updated at 30 m in 2020. Both mapping studies use the African Soil database composed of legacy digitized soil profiles. The second study of 2020 includes little additional soil profiles, but extends the number of generated soil maps produced. These studies are devoted to provide soil data (soil property maps) indirectly, but without new data collection, the connection of these maps to the reality may well be becoming fainter. It is not clear whether the available legacy data and the rate of primary data production can support the considerable societal demand for digital soil information. In this way, digital soil mapping is becoming a victim of its own success. We need to develop ratios and standards of input to output data.

## 5 | THE IMMEDIATE FUTURE

### 5.1 | Machine learning and natural language processing

Although machine learning has been exploited considerably, the continuous development of digital soil databases and computational power in the near future is likely to support a further increase in the use of machine learning. Machine learning is currently used mostly to build empirical relationships between physical, biological, and chemical soil data and environmental factors, and to predict these soil data from the pattern found in the data (see also Section 2.3). It is also widely used in chemometrics (Section 2.2), to build empirical relationships between laboratory-measured soil property values and spectra from optical sensors composed of several thousands of wavelengths. Future research will use these complex statistical and algorithmic tools in combination with pedological expertise and tacit knowledge learned through experience. This can be done in several ways—for example, by (a) including pedological knowledge in the machine learning modeling approach, or (b) extracting insights and hypotheses from the empirical relationships found in the digital soil data. Machine learning will thus be complementary to existing mechanistic models, instead of supplanting them. Machine learning models have the advantage of being highly flexible, often more accurate than mechanistic models, and to allow modeling when little is known about the process under study (Ma et al., 2019). Conversely, mechanistic models could provide physical constraints to machine learning models and guide the model in area of evident extrapolation (Follain et al., 2006). In parallel, machine learning will also be used to tackle some of the pressing challenges of digital soil science—for example, to incorporate different data sources, such as soil measurements from different sensors and laboratory techniques, or to combine environmental information from different data providers. The near future will also see the development of natural language processing in soil science. Soil science benefits from a large collection of historical (qualitative) descriptive data that are currently being digitized into digital data in text form (natural language). Although this type of information is usually disregarded in existing electronic databases, it will be possible to use the descriptive soil data to complement common numerical analysis of digital data (as in Furey et al., 2019, for example).

### 5.2 | Internet of Things (IoT) sensors

The list of digital sensors used in soil science is very broad, and most of them have been considerably exploited (e.g., volumetric soil moisture sensors, visible and near-infrared [vis-NIR] spectrometers, and airborne LIDAR). For each of these sensors, the digital data are still collected by a user. The near future will see an increase in connected objects within the concept of Internet of Things (IoT; Gubbi et al., 2013) in which sensors will “talk” to each other through the internet and wireless connections. Instead of a user collecting data from soil sensors, the digital data will be transmitted directly to the cloud where they can be processed and visualized. To date, the main obstacle in the development of IoT sensors was the cost of initial investment in wireless and connected sensors and the lack of broadband coverage in many areas of the world (Ojha et al., 2015). Cost in connected sensors has been considerably reduced with the extensive adoption of these sensors for digital and precision agriculture (Khanna & Kaur, 2019). There has also been a recent improvement in communication protocols requiring low bandwidth for areas where broadband coverage is lacking. Data also can be stored and transmitted intermittently to the cloud with satellite communications. Another obstacle, perhaps secondary in soil science compared with, for example, digital agriculture, was the fear over security issues for data privacy and ownership, and cyber attacks on data storage and exchange in the cloud. In this too, considerable progress has been made because almost every institution has now adopted license policies on data (e.g., Creative Commons licenses; Kim, 2007) and security strategies and protocol to protect against online threats. All in all, IoT sensors will permit a better spatial and temporal coverage of the soil with automated data collection and transmission, which might bring considerable advances in the understanding of short- and long-term soil variation. To achieve this, connected sensors will need to be deployed and maintained over long periods of time to obtain a quasiexhaustive picture of the variation of soil properties over a range of environmental conditions.

## 5.3 | Robotic measurements

Robotic measurement is being made operational for a wide range of soil science applications. Robots are a precision and autonomous technology that have already been widely used in digital agriculture—for instance, as automatic and mobile feeding systems that distribute fodder through the day, or as outdoor weeding machines that physically destroy the weeds using built-in high-precision GPS guidance, a series of sensors such as cameras, and statistical algorithms coupled with machine vision (Bellon-Maurel & Huyghe, 2017). As in digital agriculture, robots in soil science will enable repeated and precise measurements, coupled with sensors for in-field soil analysis. This will considerably reduce the environmental impact and the costs of future soil surveys, and increase the accuracy of the measurement protocol. At the time of writing, startups are developing autonomous robots to collect soil samples. One of them is equipped with high-precision GPS and high-speed hydraulic auger. The soil auger robot is less heavy than conventional mechanical soil auger mounted on a vehicle and has a precision of few centimeters for revisiting previous sampling locations. This obviously appeals soil scientists interested to increase the spatial and temporal coverage of soil surveys. Other robots in development are coupled with penetrometers to measure soil compaction, or electrical resistivity. Research in this direction is still preliminary, published in engineering journals and developed in collaboration between research institutions and industrial partners. No doubt that robotic soil measurements, in combination with the IoT, will soon be of valuable help in soil survey. Autonomous robots will collect soil profile cores, analyze them onsite using a set of optical sensors (e.g., vis-NIR spectrometer), and send digital information to the cloud via a set of wireless sensor nodes where the remote user will monitor the soil data collection. Adaptive sampling techniques using similar technologies are already being deployed (see, for example, Fentanes et al., 2018).

## 5.4 | Big data and mega-computation

We can expect a dramatic increase in digital data production which will trigger tremendous computational challenges. The computational power that will be needed outstrips by far the current power available in desktop and cluster computers. Soil scientists will need to develop computational tools, software packages, and pipelines for accessing and analyzing the data. In addition to requiring advanced training in computer science, digital soil data analysis is likely to be performed in large research institutions. These benefit from the most specialized researchers and social advantages of acquaintance with researchers with specific expertise. The same institutions will possess the largest soil data generation tools (e.g., remote sensing imagery) and the facilities to analyze them. Only well-financed and centralized institutions will pay for the expensive computer power and will maintain specialized infrastructures. Still, analysis of soil "big data" in a distributed fashion is likely to increase. Privately owned cloud computing platforms have massive, virtually unlimited capacity and are cost efficient for individual researchers or small institutions. The Africa Soil Information Service (AfSIS) Soil Chemistry database, for example, is currently stored in the Amazon Web Service (AWS), and tens of terrabytes of Sentinel data are preprocessed in the cloud using AWS. Strasser and Edwards (2017) contend that "big data" are too large to be fully exploited by the institutions that produced them. They are also increasingly made publicly available. The availability of open large soil datasets and cheap distributed computing will inevitably change to landscape of science, where anyone could potentially perform research outside large research institutions. All in all, in the near future, mega-computing and big data will increasingly be important in digital soil science, but there will be no restrictions on what can be computed. Models built on these big data will become more complicated than today's models but will also include more variables and processes and predict at finer time steps and spatial resolutions. Models will always be at the forefront of the available computational resources (Rossiter, 2018).

## 5.5 | Global soil understanding

There is an increasing understanding that the soil resource is finite. We need to monitor the state of the soil globally to understand what role the soil plays in earth system functioning, the drivers of soil dynamics, and to unravel the major natural and artificial processes (Pennock et al., 2015). We also need to decide what kind of intervention we can perform locally to secure and enhance the soil. Most digital tools are well adapted (i.e., "scalable") to support studies at a global scale, and we speculate that in the near future, the global understanding of soils will increase jointly with development of digital data availability. We will know the properties and functions of soil at very local scales everywhere on the planet. We will know how this local functioning scales up to watershed and ecoregions. To achieve this, we will likely make more use of remote sensing imagery and spatially continuous measurements of soil properties made by various sensors. Data will not only be composed of digits, but also digitized (analog) qualitative data from past field soil descriptions. In fact, soil scientists have already started to compile global digital data for understanding the soils. The digital soil maps of the world (Hengl et al., 2017), for example, were a "proof of concept" that global digital soil data can be compiled, harmonized, processed, and distributed digitally through an online platform. Despite the drawbacks of such global maps for regional or local scale soil understanding (Mulder et al., 2016), these efforts foreshadow many others that are already taking place or are expected to appear in the coming years. The global maps of earthworm diversity (Phillips et al., 2019) resulting from a global effort in data compilation, for example, have revealed that climatic drivers are more likely to explain earthworms diversity than soil types. Corollary to the development in global digital soil data sharing, the digital techniques to handle these dataset will also become more complex. The IoT will enable continuous feedback between measurements and soil models. Modeling of soil will be available in near real time using cloud computing and sensor networks. Finally, global models monitoring the state and evolution of the soil will be constantly refined and updated at local scale.

## 6 | BEYOND DIGITAL: A LOOK INTO THE CRYSTAL BALL

Predicting the impact of the digital in soil science is limited by our current knowledge on the forces that will drive soil science research in the far future. We ignore which new ideas, pressing challenges, technologies, techniques, and sensors will transform our discipline. Thus, in predicting how digital soil science may develop in a few decades to a century, there is always the barrier of unpredictability and high uncertainty that any long-term prediction involves. Instead of making predictions, we speculate on three scenarios that may best represent our intention to project potential future events. In each of the scenarios, we hypothesize that the digital convergence will be a driving factor of soil science research (Figure 5), perhaps even more so than in present-day soil science.

### 6.1 | Digitally enabled regenerative soil science

In this scenario, we consider the digital soil science applied to solving the pressing problems of degraded soils and the environmental issues that accompany this degradation. The rate of world population growth will trigger a doubling of food production in this century (Richter et al., 2007). This will put unprecedented pressure on soils. We are currently losing the best agricultural soils due to urbanization, salinization, and erosion, and in parallel, the new soils put into production are often of lower quality and more prone to quick degradation (Kopittke et al., 2019). If the current pattern of soil use continues, it is expected that soils will overshoot their carrying capacity for producing a set of ecosystem services. The human population is a great consumer of soil resources, but the quantity of soils available globally is finite. Simultaneously to soil resources becoming scarce, future growth in food production will also depend intimately on intensification and productivity gains.

We speculate that in the future, the digital convergence will support regenerative soil science with two main aspects: (a) preservation of soils to support an optimal use, and (b) regeneration to enhance soils to a desired state. In each of these cases, the ambition of regenerative soil science is to regenerate and preserve the quality of soil functioning and to improve the capacity of providing a series of ecosystem services, in particular biomass and food production. A set integrated and interconnected digital tools will measure and sense ecosystem and soil functioning and provide real-time proxy information on soil condition. For example, bird and insect measurements can be used as proxy of aboveground biodiversity, soil carbon measurement can be used as indicator of the overall soil capability to supporting crop production, soil biodiversity abundance and diversity can be sensed for estimating belowground soil biodiversity, remote sensing will be monitoring climatic conditions and vegetation production, and so on. All these digital data integrated into one interconnected system will serve the purpose of constant and multiscale monitoring of the soil resource at a site.

It is not just that the soil is monitored, but that it is steered towards a desired state by optimizing all the components of the ecosystem for an intended use (i.e., preservation or regeneration): fallow periods can be reduced, and soil conditions can be optimized for specific soil microbes that will preclude invasive pests in agroecosystems. In this optimal use of the soil, artificial inputs such as irrigation, chemical fertilizers, and pesticides are bypassed by efficient soil management

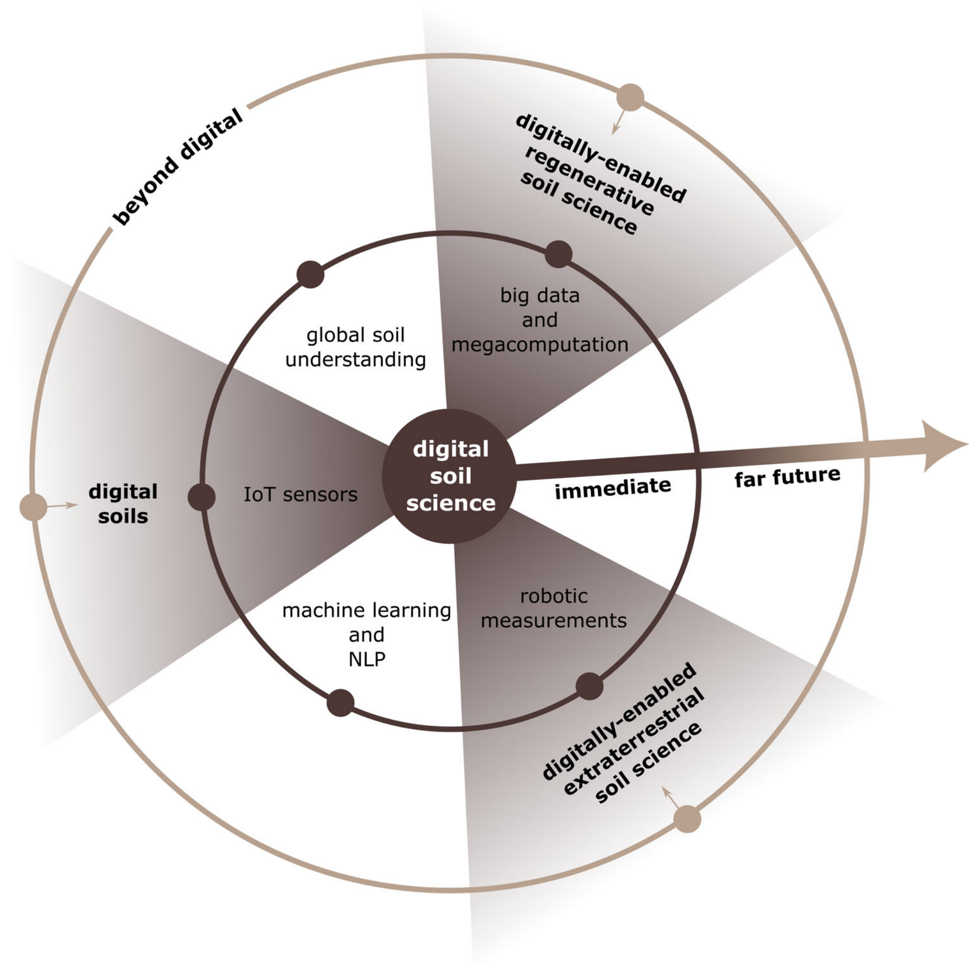


**FIGURE 5** Summary diagram of the future aspects of soil science enabled by the digital. The immediate future of digital soil science supports the development of digital technologies and techniques applied to the study of the soil, as well as an increasing global understanding of soils. These new techniques and technologies can in turn permit to go beyond digital in the far future, to support the creation of digital soils (i.e., a digital twin of the soil), and assist regenerative and extraterrestrial soil science. IoT, Internet of Things; NLP, natural language processing

strategies. In short, the digital will enable the production of soils, optimal for a specific human and ecosystem use. Well-drained and fertile soils for food production, soils providing shelters for microorganism diversity, and even clayey soils for the production of pottery. It is in fact one step towards soil reengineering (i.e., manipulation of the soil to cancel out the most harmful human misuses—for example, to steer soil towards a desirable state of optimal chemical capture of carbon).

A fact that is immediately obvious is that not all soils are of similar capability or can be sufficiently improved for a specific propose (e.g., for crop production). There will be the need for a digital and autonomous “governing entity” (in other words, a pedological “big brother”), which could perhaps rely on the block-chain technology (an interconnected set of databases storing the information). This entity will process the constant stream of digital data from a network of soil sensors to determine which purpose best suits a soil (optimal allocation) while considering its current condition. The entity could also preclude the connection of local food production with distant markets.

## 6.2 | Digitally enabled extraterrestrial soil science

In this scenario, we consider that the new frontier in soil science will be the study of extraterrestrial soils. The term extraterrestrial soils is used here to describe the regolith found in planetary bodies other than the Earth. They have some similarities with Earth’s soils in that they result from the physical and chemical degradation of the parent rock over time, but unlike the Earth’s soils none of the known extraterrestrial soils show evidence of a biological component or distinctive "pedological" horizons along profiles (Cameron, 1963). In fact, it is likely that the soil forming factors important for Earth soils (e.g., vegetation or climate) are nonexistent or negligible in extraterrestrial soils, where soil forming factors such as topography may be predominant.

In the short term, say decades to a century, exploration and understanding of extraterrestrial soils will essentially be digital and rely on inferred rather than direct measurements. We already have access to lunar and planetary surface samples—for example, brought by Apollo 11 in 1969 or in the close future in 2031 with the sample-return mission from Mars. These samples can be analyzed but constitute a very limited sample of the potential variability of extraterrestrial soils. We also have limited access to a range of background properties and characteristic extraterrestrial ecosystems in which these soils form. This is because there are many factors that control the basic physical and chemical characteristics of the soils in the extraterrestrial environment that are unknown or difficult to measure (e.g., gas exchange or pH; Certini et al., 2009). Current methods for field data description are inadequate for extraterrestrial soil exploration (see also Section 3). Soil pH, for example, is difficult to estimate or infer remotely.

The close future of soil science will develop an extraterrestrial science in which inferring the soil data, rather than analyzing physical soil material, will be the rule and in which understanding of extraterrestrial soil will rely almost entirely on the digital, to the point where accessing the soil material will be the exception. Extraterrestrial soil scientists will have no intimate or direct contact with the soils. In fact, several of these elements are already in place. Past missions on Mars have already relied on inferred measurements, using gas chromatography and mass spectrometry, and new sensors will soon be deployed. For example, an automated microchip electrophoresis analyzer mounted on a rover could detect organic biosignatures in extraterrestrial soils (Mora et al., 2020). In the far future, humans might populate some new planetary and obtain access to physical soil material, but there will always be new planetary systems to explore, so that understanding of extraterrestrial soils will always be at the limit of what the digital enables.

### 6.3 | From digital soil science to digital soils

In the first scenario, we assumed that soils in the future will remain a precious resource and that the digital will serve the purpose of managing and regenerating this resource to a desired state. This scenario, conversely, assumes that in the far future the digital would be powerful enough to substitute for the soils resource as we currently know it. This scenario is inspired by Haff (2014), who imagined a far future in which the soils, rivers, and biology in a technology-dominated planet are considered as technological artifacts rather than natural systems.

Soils would not disappear completely; the soil physical material stays in place but is populated by microsensors, microcomputers, and micro-actuators that determine its dynamics. We have shown previously (Section 2) that with rapid technological change, sensors are becoming smaller, faster, and multifunctional. Barometers, light, accelerometers, and gyroscopes are already present in ordinary watches, connected physical activity watches, coffee mugs, or domestic appliances and have been translated into sensors for environmental monitoring. Many elements of the environment are monitored by sensors, to the point where the environment is computerized and can be controlled. Synthetic environments are being created for food production, under a digitally controlled environment where temperature, light, and humidity are regulated by algorithms and sensing devices to ensure optimal growth conditions. The population of the soil material with large number of microsensors coupled with wireless technologies and the IoT may also end up with the creation of a digital twin of the soil (as briefly discussed in Searle et al., 2021), or, as Haff (2014) puts it, a computerized soil.

The digital twin of the soil results from deliberately scattering a large number of grain-size computers, sensors, and actuators over a land area. Power is provided by ambient energy, mostly solar but also wind or seismic vibration. Individual microcomputers are networked one to another and wirelessly connected to a central processing system tasked to receive and analyze the constant stream of data coming from the land. Microsensors would constantly measure dynamic soil properties, such as moisture, temperature, water holding capacity, or structure and provide a real-time picture of the state of the soil. The measurement of these sensors is coordinated with the centralized processing system, which may instruct a response if the state of the soil becomes unsuitable. Local actuators would be activated to perform actions and apply forces on the soil. Haff (2014) shows the example of actuators counteracting the effect of erosion, but more complex systems can be imagined, involving chemical and biological responses of the soil (e.g., to chemically sequester carbon). Meanwhile, the sensors send a constant feedback on the soil properties to the central processing system. The soil properties are integrated into a computer-based hierarchy of controls. Soil in this sense become a programmable medium divorced from the classical natural influence, and independent of nature-based solution for its management.

## 7 | CONCLUSION

- Digital convergence has changed soil science and continues to do so.

- There have been major successes via remote and proximal soil sensing, digital soil mapping, soil biogenomics, and cell phone applications.
- There have been notable failures such as the relative lack of data federation and sharing, truly digital field soil description, and development of digital taxonomic systems.
- There are potential pitfalls with digital convergence including lack of new soil theory, lack of soil science knowledge by researchers, and trying to do too much with too little information or data.
- In the immediate future, digital soil science will be dominated by machine learning, the IoT, robotic measurement, and big data, all leading to better global soil understanding.
- In the far future, digital soil science may allow use to regenerate soil here on earth, and create multifunctional soils on other planets and potentially create intelligent self-organizing goal-oriented soils.

## AUTHOR CONTRIBUTIONS

Alexandre M. J.-C. Wadoux: Conceptualization; Investigation; Writing-original draft; Writing-review & editing. Alex B. McBratney: Conceptualization; Writing-original draft; Writing-review & editing.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ORCID

*Alexandre M. J.-C. Wadoux* https://orcid.org/0000-0001-7325-9716

## REFERENCES


Aitkenhead, M., Donnelly, D., Coull, M., & Gwatkin, R. (2016). Estimating soil properties with a mobile phone. In A. E. Hartemink & B. Minasny (Eds.), *Digital soil morphometrics* (pp. 89–110). Springer.

Appel, M., & Pebesma, E. (2019). On-demand processing of data cubes from satellite image collections with the gdalcubes library. *Data*, *4*, 92. https://doi.org/10.3390/data4030092

Basher, L. R. (1997). Is pedology dead and buried? *Soil Research*, *35*, 979–994. https://doi.org/10.1071/S96110

Beckett, P. H. T., Webster, R., Mcneil, G. M., & Mitchell, C. W. (1972). Terrain evaluation by means of a data bank. *Geographical Journal*, *138*, 430–449. https://doi.org/10.2307/1795497

Bellon-Maurel, V., & Huyghe, C. (2017). Putting agricultural equipment and digital technologies at the cutting edge of agroecology. *Oléagineux, Corps Gras, Lipides*, *24*, 1–7.

Biggs, A., Holz, G., McKenzie, D., Doyle, R., & Cattle, S. (2018). Pedology: A vanishing skill in Australia? *Soil Science Policy Journal*, *1*, 24–35.

Bouma, J. (2010). Implications of the knowledge paradox for soil science. *Advances in Agronomy*, *106*, 143–171. https://doi.org/10.1016/S0065-2113(10)06004-9

Bouma, J., Stoorvogel, J., Van Alphen, B. J., & Booltink, H. W. G. (1999). Pedology, precision agriculture, and the changing paradigm of agricultural research. *Soil Science Society of America Journal*, *63*, 1763–1768. https://doi.org/10.2136/sssaj1999.6361763x

Bui, E. N., & Moran, C. J. (2003). A strategy to fill gaps in soil survey over large spatial extents: An example from the Murray–Darling basin of Australia. *Geoderma*, *111*, 21–44. https://doi.org/10.1016/S0016-7061(02)00238-0

Buringh, P. (1954). The analysis and interpretation of aerial photographs in soil survey and land classification. *NJAS Wageningen Journal of Life Sciences*, *2*, 16–26.

Bushnell, T. M. (1929). Aerial photography and soil survey. *Soil Science Society of America Journal*, *10*, 23–28. https://doi.org/10.2136/sssaj1929.036159950B1020010004x

Cameron, R. (1963). The role of soil science in space exploration. *Space Science Reviews*, *2*, 297–312. https://doi.org/10.1007/BF00216782

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., Mcdonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*, 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, *6*, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Certini, G., Scalenghe, R., & Amundson, R. (2009). A view of extraterrestrial soils. *European Journal of Soil Science*, *60*, 1078–1092. https://doi.org/10.1111/j.1365-2389.2009.01173.x

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., McBratney, A. B., Wood, E. F., & Yimam, Y. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. *Water Resources Research*, *55*, 2916–2938. https://doi.org/10.1029/2018WR022797

Dehaan, R., & Taylor, G. R. (2003). Image-derived spectral endmembers as indicators of salinisation. *International Journal of Remote Sensing*, *24*, 775–794. https://doi.org/10.1080/01431160110107635

Fentanes, J. P., Gould, I., Duckett, T., Pearson, S., & Cielniak, G. (2018). 3-D soil compaction mapping through kriging-based exploration with a mobile robot. *IEEE Robotics and Automation Letters*, *3*, 3066–3072. https://doi.org/10.1109/LRA.2018.2849567

Flynn, K. D., Bagnall, D. K., & Morgan, C. L. S. (2020). Evaluation of SLAKES, a smartphone application for quantifying aggregate stability, in high-clay soils. *Soil Science Society of America Journal*, *84*, 345–353. https://doi.org/10.1002/saj2.20012

Follain, S., Minasny, B., McBratney, A. B., & Walter, C. (2006). Simulation of soil thickness evolution in a complex agricultural landscape at fine spatial and temporal scales. *Geoderma*, *133*, 71–86. https://doi.org/10.1016/j.geoderma.2006.03.038

Furey, J., Davis, A., & Seiter-Moser, J. (2019). Natural language indexing for pedoinformatics. *Geoderma*, *334*, 49–54. https://doi.org/10.1016/j.geoderma.2018.07.050

Golicz, K., Hallett, S., Sakrabani, R., & Ghosh, J. (2020). Adapting smartphone app used in water testing, for soil nutrient analysis. *Computers and Electronics in Agriculture*, *175*, 105532. https://doi.org/10.1016/j.compag.2020.105532

Gómez-Robledo, L., López-Ruiz, N., Melgosa, M., Palma, A. J., Capitán-Vallvey, L. F., & Sánchez-Marañón, M. (2013). Using the mobile phone as Munsell soil-colour sensor: An experiment under controlled illumination conditions. *Computers and Electronics in Agriculture*, *99*, 200–208. https://doi.org/10.1016/j.compag.2013.10.002

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*, 1645–1660. https://doi.org/10.1016/j.future.2013.01.010

Haff, P. K. (2014). The far future of soil. In G. J. Churchman & E. R. Landa (Eds.), *The soil underfoot: Infinite possibilities for a finite resource* (pp. 61–72). CRC Press.

Haines, W. B., & Keen, B. A. (1925). Studies in soil cultivation. II. A test of soil uniformity by means of dynamometer and plough. *The Journal of Agricultural Science*, *15*, 387–394. https://doi.org/10.1017/S0021859600006821

Han, P., Dong, D., Zhao, X., Jiao, L., & Lang, Y. (2016). A smartphone-based soil color sensor: For soil type classification. *Computers and Electronics in Agriculture*, *123*, 232–241. https://doi.org/10.1016/j.compag.2016.02.024

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., & Schröder, B. (2012). Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma*, *185*, 37–47. https://doi.org/10.1016/j.geoderma.2012.04.001

Hartemink, A. E., & Minasny, B. (2014). Towards digital soil morphometrics. *Geoderma*, *230*, 305–317. https://doi.org/10.1016/j.geoderma.2014.03.008

Hengl, T., & Mendes de Jesus, J. (2015). SoilInfo App: Global soil information on your palm. In *EGU General Assembly Conference Abstracts* (p. 2372). European Geosciences Union.

Hengl, T., Mendes De Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., Macmillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hengl, T., Wheeler, I., & MacMillan, R. A. (2018). A brief introduction to open data, open source software and collective intelligence for environmental data creators and users. *Peer J Preprints*, *6*, e27127v2.

Hudson, B. D. (1992). The soil survey as paradigm-based science. *Soil Science Society of America Journal*, *56*, 836–841. https://doi.org/10.2136/sssaj1992.03615995005600030027x

Hughes, P., McBratney, A., Huang, J., Minasny, B., Hempel, J., Palmer, D. J., & Micheli, E. (2017). Creating a novel comprehensive soil classification system by sequentially adding taxa from existing systems. *Geoderma Regional*, *11*, 123–140. https://doi.org/10.1016/j.geodrs.2017.10.004

Ibanez, J. J. (2006). Future of soil science. In (Ed.), *Future of soil science* (pp. 60–62). International Union of Soil Sciences.

Jahn, R., Blume, H. P., Asio, V. B., Spaargaren, O., & Schad, P. (2006). *Guidelines for soil description*. FAO.

Jian, J., Du, X., & Stewart, R. D. (2020). A database for global soil health assessment. *Scientific Data*, *7*, 1–8. https://doi.org/10.1038/s41597-020-0356-3

Jones, E. J., & McBratney, A. B. (2016). What is digital soil morphometrics and where might it be going? In A. E. Hartemink & B. Minasny (Eds.), *Digital soil morphometrics* (pp. 1–15). Springer.

Kelly, J., Häggblom, M., & Tate, III, R. L. (1999). Effects of the land application of sewage sludge on soil heavy metal concentrations and soil microbial communities. *Soil Biology and Biochemistry*, *31*, 1467–1470. https://doi.org/10.1016/S0038-0717(99)00060-7

Kempen, B., Brus, D. J., Heuvelink, G. B. M., & Stoorvogel, J. J. (2009). Updating the 1: 50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma*, *151*, 311–326. https://doi.org/10.1016/j.geoderma.2009.04.023

Khanna, A., & Kaur, S. (2019). Evolution of internet of Things (IoT) and its significant impact in the field of precision agriculture. *Computers and Electronics in Agriculture*, *157*, 218–231. https://doi.org/10.1016/j.compag.2018.12.039

Kim, M. (2007). The Creative Commons and copyright protection in the digital era: Uses of Creative Commons licenses. *Journal of Computer-Mediated Communication*, *13*, 187–209. https://doi.org/10.1111/j.1083-6101.2007.00392.x

Kopittke, P. M., Menzies, N. W., Wang, P., Mckenna, B. A., & Lombi, E. (2019). Soil and the intensification of agriculture for global food security. *Environment International*, *132*, 105078. https://doi.org/10.1016/j.envint.2019.105078

Kristof, S. J., Baumgardner, M. F., & Johannsen, C. J. (1973). *Spectral mapping of soil organic matter* [LARS Technical Reports]. Purdue University.

Lagacherie, P., Legros, J. P., & Burfough, P. A. (1995). A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. *Geoderma*, *65*, 283–301. https://doi.org/10.1016/0016-7061(94)00040-H

Lagacherie, P., & McBratney, A. B. (2006). Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. In P. Lagacherie, A. B. McBratney, & M. Voltz (Eds.), *Developments in soil science* (pp. 3–22). Elsevier.

Lakeman-Fraser, P., Gosling, L., Moffat, A. J., West, S. E., Fradera, R., Davies, L., Ayamba, M. A., & Van Der Wal, R. (2016). To have your citizen science cake and eat it? delivering research and outreach through Open Air Laboratories (OPAL). *BMC Ecology*, *16*, 57–70. https://doi.org/10.1186/s12898-016-0065-0

Legros, J.-P. (2006). *Mapping of the soil*. Science Publishers.

Levin, N., Ben-Dor, E., & Singer, A. (2005). A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments. *International Journal of Remote Sensing*, *26*, 5475–5492. https://doi.org/10.1080/01431160500099444

Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., … Wang, L.-W. (2017). The australian geoscience data cube: Foundations and lessons learned. *Remote Sensing of Environment*, *202*, 276–292. https://doi.org/10.1016/j.rse.2017.03.015

Lobry de Bruyn, L., Jenkins, A., & Samson-Liebig, S. (2017). Lessons learnt: Sharing soil knowledge to improve land management and sustainable soil use. *Soil Science Society of America Journal*, *81*, 427–438. https://doi.org/10.2136/sssaj2016.12.0403

Ma, Y., Minasny, B., Malone, B. P., & McBratney, A. B. (2019). Pedology and digital soil mapping (DSM). *European Journal of Soil Science*, *70*, 216–235. https://doi.org/10.1111/ejss.12790

McBratney, A. B., & Minasny, B. (2010). The sun has shone here antecedently. In R. A. Viscarra Rossel, A. B. McBratney, & B. Minasny (Eds.), *Proximal soil sensing* (pp. 67–75). Springer.

McBratney, A. B., Minasny, B., Cattle, S. R., & Vervoort, R. W. (2002). From pedotransfer functions to soil inference systems. *Geoderma*, *109*, 41–73. https://doi.org/10.1016/S0016-7061(02)00139-8

McBratney, A. B., & Webster, R. (1981). Spatial dependence and classification of the soil along a transect in northeast Scotland. *Geoderma*, *26*, 63–82. https://doi.org/10.1016/0016-7061(81)90076-8

McKenzie, N. J., Grundy, M. J., Webster, R., & Ringrose-Voase, A. J. (2008). *Guidelines for surveying soil and land resources*. CSIRO Publishing.

Minasny, B., & McBratney, A. B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma*, *264*, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Moore, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, *38*, 114–117.

Mora, M. F., Kehl, F., Tavares Da Costa, E., Bramall, N., & Willis, P. A. (2020). Fully automated microchip electrophoresis analyzer for potential life detection missions. *Analytical Chemistry*, *92*, 12959–12966. https://doi.org/10.1021/acs.analchem.0c01628

Muir, J. W., & Hardie, H. G. M. (1962). A punched-card system for soil profiles. *Journal of Soil Science*, *13*, 249–253. https://doi.org/10.1111/j.1365-2389.1962.tb00704.x

Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping: A review. *Geoderma*, *162*, 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Mulder, V. L., Lacoste, M., Richer-De-Forges, A. C., Martin, M. P., & Arrouays, D. (2016). National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, *263*, 16–34. https://doi.org/10.1016/j.geoderma.2015.08.035

Ojha, T., Misra, S., & Raghuwanshi, N. S. (2015). Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges. *Computers and Electronics in Agriculture*, *118*, 66–84. https://doi.org/10.1016/j.compag.2015.08.011

Orgiazzi, A., Dunbar, M. B., Panagos, P., De Groot, G. A., & Lemanceau, P. (2015). Soil biodiversity and DNA barcodes: Opportunities and challenges. *Soil Biology and Biochemistry*, *80*, 244–250. https://doi.org/10.1016/j.soilbio.2014.10.014

Padarian, J., & McBratney, A. B. (2020). A new model for intra-and inter-institutional soil data sharing. *SOIL*, *6*, 89–94. https://doi.org/10.5194/soil-6-89-2020

Palacios-Orueta, A., & Ustin, S. L. (1998). Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis. *Remote Sensing of Environment*, *65*, 170–183. https://doi.org/10.1016/S0034-4257(98)00024-8

Pennock, D., McKenzie, N., & Montanarella, L. (2015). *Status of the world's soil resources*. FAO.

Philip, J. R. (1991). Soils, natural science, and models. *Soil Science*, *151*, 91–98. https://doi.org/10.1097/00010694-199101000-00011

Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., Ferlian, O., Gongalsky, K. B., Van Den Hoogen, J., Krebs, J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E. M., Bennett, J. M., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., … Eisenhauer, N. (2019). Global distribution of earthworm diversity. *Science*, *366*, 480–485. https://doi.org/10.1126/science.aax4851

Ragg, J. M. (1977). The recording and organization of soil field data for computer areal mapping. *Geoderma*, *19*, 81–89. https://doi.org/10.1016/0016-7061(77)90016-7

Richter, D. D., Hofmockel, M., Callaham, M. A., Powlson, D. S., & Smith, P. (2007). Long-term soil experiments: Keys to managing Earth's rapidly changing ecosystems. *Soil Science Society of America Journal*, *71*, 266–279. https://doi.org/10.2136/sssaj2006.0181

Robinson, N. J., Dahlhaus, P. G., Wong, M., Macleod, A., Jones, D., & Nicholson, C. (2019). Testing the public–private soil data and information sharing model for sustainable soil management outcomes. *Soil Use and Management*, *35*, 94–-104. https://doi.org/10.1111/sum.12472

Rogers, R. H., Wilson, C. L., Reed, L. E., Shah, N. J., Akeley, R., Mara, T. G., & Smith, V. E. (1975). Environmental monitoring from spacecraft data. In C. D. McGillem & D. Morrison (Eds.), *Symposium on machine processing of remotely sensed data* (p. 76). Purdue University.

Rossiter, D. G. (2018). Past, present & future of information technology in pedometrics. *Geoderma*, *324*, 131–137.

Rudeforth, C. C. (1975). Storing and processing data for soil and land use capability surveys. *Journal of Soil Science*, *26*, 155–168. https://doi.org/10.1111/j.1365-2389.1975.tb01940.x

Rust, W. M. (1938). A historical review of electrical prospecting methods. *Geophysics*, *3*, 1–6. https://doi.org/10.1190/1.1439461

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Opensource, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schlumberger, C. (1920). *Étude sur la Prospection Électrique du Soussol*. Gauthier-Villars.

Schmugge, T. (1978). Remote sensing of surface soil moisture. *Journal of Applied Meteorology*, *17*, 1549–1557. https://doi.org/10.1175/1520-0450(1978)017%3c1549:RSOSSM%3e2.0.CO;2

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., … Andrews, K. (2021). Digital soil mapping and assessment for Australia and beyond: A propitious future. *Geoderma Regional*, *24*, e00359. https://doi.org/10.1016/j.geodrs.2021.e00359

Shelley, W. (2012). *mySoil app for smartphones*. British Geological Survey. https://www.bgs.ac.uk/technologies/apps/mysoil-app/

Shonk, J. L., Gaultney, L. D., Schulze, D. G., & Van Scoyoc, G. E. (1991). Spectroscopic sensing of soil organic matter content. *Transactions of the American Society of Agricultural and Biological Engineers*, *34*, 1978–1984. https://doi.org/10.13031/2013.31826

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*, *107*, 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7

Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., & Sharp, J. (2016). Evaluation of an inexpensive sensor to measure soil color. *Computers and Electronics in Agriculture*, *121*, 141–148. https://doi.org/10.1016/j.compag.2015.11.014

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., De Brogniez, D., Montanarella, L., Hong, S. Y., Rawlins, B. G., & Field, D. J. (2015). Global soil organic carbon assessment. *Global Food Security*, *6*, 9–16. https://doi.org/10.1016/j.gfs.2015.07.001

Strasser, B. J., & Edwards, P. N. (2017). Big data is the answer… But what is the question? *Osiris*, *32*, 328–345. https://doi.org/10.1086/694223

Tate III, R. L. (2020). *Soil microbiology*. (3rd ed.). Wiley-Blackwell.

Taylor, J. A., McBratney, A. B., Viscarra Rossel, R. A., Minasny, B., Taylor, H. J., Whelan, B. M., & Short, M. (2006). *Development of a multi-sensor platform for proximal soil sensing* [Presentation 56-15]. 18th World Congress of Soil Science July, Philadelphia, PA.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE*, *6*, e21101. https://doi.org/10.1371/journal.pone.0021101

Thomas, G. W. (1992). In defense of observations and measurements. *Soil Science Society of America Journal*, *56*, 1979–1979. https://doi.org/10.2136/sssaj1992.03615995005600060054x

Vaudour, E., Gomez, C., Fouad, Y., & Lagacherie, P. (2019). Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment*, *223*, 21–33. https://doi.org/10.1016/j.rse.2019.01.006

Viscarra Rossel, R. A., Adamchuk, V., Sudduth, K. A., McKenzie, N. J., & Lobsey, C. (2011). Proximal soil sensing: An effective approach for soil measurements in space and time. *Advances in Agronomy*, *113*, 243–291. https://doi.org/10.1016/B978-0-12-386473-4.00005-1

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., … Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, *155*, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., Van Elsas, J. D., Bailey, M. J., Nalin, R., & Philippot, L. (2009). TerraGenome: A consortium for the sequencing of a soil metagenome. *Nature Reviews Microbiology*, *7*, 252–252. https://doi.org/10.1038/nrmicro2119

Wadoux, A. M. J.-C., Malone, B., Minasny, B., Fajardo, M., & McBratney, A. B. (2021). *Soil spectral inference with R: Analysing digital soil spectra with the R programming environment*. Progress in Soil Science, Springer.

Wadoux, A. M. J.-C., & McBratney, A. B. (2021). Hypotheses, machine learning and soil mapping. *Geoderma*, *383*, 114725. https://doi.org/10.1016/j.geoderma.2020.114725

Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wadoux, A. M. J.-C., Román-Dobarco, M., & McBratney, A. B. (2021). Perspectives on data-driven soil research. *European Journal of Soil Science*, *72*, 1675–1689. https://doi.org/10.1111/ejss.13071

Webster, R., & Beckett, P. H. (1970). Terrain classification and evaluation using air photography: A review of recent work at Oxford. *Photogrammetria*, *26*, 51–75. https://doi.org/10.1016/0031-8663(70)90037-2

Webster, R., & Burrough, P. A. (1972). Computer-based soil mapping of small areas from sample data: II. Classification smoothing. *Journal of Soil Science*, *23*, 222–234. https://doi.org/10.1111/j.1365-2389.1972.tb01655.x

White, J. L. (1971). Interpretation of infrared spectra of soil minerals. *Soil Science*, *112*, 22–31. https://doi.org/10.1097/00010694-197107000-00005

Whiteley, A. S., Jenkins, S., Waite, I., Kresoje, N., Payne, H., Mullan, B., Allcock, R., & O'donnell, A. (2012). Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *Journal of Microbiological Methods*, *91*, 80–88. https://doi.org/10.1016/j.mimet.2012.07.008

Wischmeier, W. H. (1955). Punched cards record runoff and soil-loss data. *Agricultural Engineering*, *36*, 664–666.

Wulf, H., Mulder, T., Schaepman, M. E., Keller, A., & Jörg, P. C. (2015). *Remote sensing of soils*. University of Zurich, Remote Sensing Laboratories.

Xu, C., Du, X., Yan, Z., & Fan, X. (2020). ScienceEarth: A big data platform for remote sensing data processing. *Remote Sensing*, *12*, 607. https://doi.org/10.3390/rs12040607

Yu, L., & Gong, P. (2012). Google Earth as a virtual globe tool for Earth science applications at the global scale: Progress and perspectives. *International Journal of Remote Sensing*, *33*, 3966–3986. https://doi.org/10.1080/01431161.2011.636081

**How to cite this article:** Wadoux AMJ-C, McBratney AB. Digital soil science and beyond. *Soil Sci Soc Am J*. 2021; 1–19. https://doi.org/10.1002/saj2.20296